Spring 2005

The Effects of Parcels and Latent Variable Scores on the Detection of Interactions in Structural Equation Modeling

Thomas D. Fletcher
Old Dominion University

THE EFFECTS OF PARCELS AND LATENT VARIABLE SCORES ON THE DETECTION OF INTERACTIONS IN STRUCTURAL EQUATION MODELING

by

Thomas D. Fletcher
M.S. December, 2002, Old Dominion University
B.S. December, 1999, Georgia State University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

OLD DOMINION UNIVERSITY
May 2005

Approved by:

Terry L. Dickinson (Co-Director)

Debra A. Major (Co-Director)

Louis H. Janda (Member)

Dayanand N. Naik (Member)

ABSTRACT

THE EFFECTS OF PARCELS AND LATENT VARIABLE SCORES ON THE DETECTION OF INTERACTIONS IN STRUCTURAL EQUATION MODELING

Thomas D. Fletcher
Old Dominion University, 2005
Co-Directors of Advisory Committee: Dr. Terry L. Dickinson
Dr. Debra A. Major

Numerous theories in the behavioral and organizational sciences involve the regression of an outcome variable on component terms and their product to evaluate interaction effects. There are numerous statistical difficulties with this multiple regression approach. The most serious is measurement error, requiring the use of structural equation modeling. Jöreskog and Yang (1996) described a nonlinear structural equation modeling procedure that incorporates mean structures in the covariance analysis. They demonstrated that only one indicator for the product term is necessary for model identification. Unfortunately, the Jöreskog-Yang procedure leads to biased estimates of the product coefficient. In this dissertation, I propose that (1) the proper use of item parcels can reduce bias in estimates, and (2) that using a relatively new technique of analysis (creation of latent variable scores) can also be fruitful in removing measurement error and improving the estimation of product terms. Two studies investigated these proposals. In Study 1, archival data were analyzed using the proposed techniques. The interaction hypothesis tested by the various techniques is that a competitive climate influences perceptions of coworker support, and that this relationship is moderated by (interacts with) a person's level of trait competitiveness. Study 2 involved a Monte Carlo investigation of methods for estimating an interaction effect. The Monte Carlo research included design factors for (a) effect size, (b) parceling strategy, and (c) method of analysis. Study 1 demonstrated that method of analysis and parceling strategy affected the detection of the moderator effect of competition on two types of coworker support (instrumental and affective). Variability in the *t*-tests and effect size indices lend credibility for the need for the Monte-Carlo investigation. Study 2 demonstrated that (1) there is greater variability in the estimation of the interaction effect with the Jöreskog-Yang method than the latent variable scores method, (2) parceling strategy has the most influence on the interaction effect in the Jöreskog-Yang method, and this effect is dependent upon which strategy is used, and (3) the latent variable score method is superior to the Jöreskog-Yang method with respect to statistical decision making (i.e., fewer Type II errors). Practical implications and future research directions are considered.

ACKNOWLEDGMENTS

This dissertation is not the product of one research study, but rather the culmination of five long years of hard work. I have encountered many individuals along the way that deserve thanks.

First, I have to acknowledge the National Science Foundation for support under Grant No. 0204430. Data used in this dissertation was collected under this grant. Further, financial support (directly and indirectly) made this dissertation possible.

Second, I have to give many thanks to Mary Boswell, Peggy Kinard, and Jackie Winston for always ensuring that the i's were dotted and the t's were crossed throughout my course of study at Old Dominion.

I thank Barbara Winstead as chair of the department for ensuring a culture that values intrinsic motivation. There were many times that she enabled me to pursue interests not readily afforded by the department or the university.

I would like to thank Glynn Coates and Terry Dickinson for directing me toward interests in quantitative psychology, knowingly or unknowingly. Their inspiration has indirectly led to this dissertation and directly to pursuing my intrinsic interests and shaping my career path.

Finally, I must give many thanks to Debbie Major for enabling my intrinsic interests in organizational psychology. She has served as my graduate school advisor for five years. In that capacity she has mentored me into a successful graduate student. She has shown me that I can pursue a career that utilizes and takes advantage of my most dominant personality trait – curiosity. She has also encouraged me to be helpful to others – a necessary trait for a successful academic.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

There are a number of theories in the behavioral and organizational sciences that require the use of statistical interactions for formal hypothesis testing. Examples include, but are not limited to contingency theories of leadership (House, 1971, 1996; Kerr & Jermier, 1978; Yukl, 2002), expectancy theory of motivation (Vroom, 1964), theory of reasoned action (Ajzen & Fishbein, 1980), and person by situation influences (Pervin, 1989; Ross & Nisbett, 1991; Schneider, 1983). More generally, interactional psychology developed in recognition that individual behavior results from an interaction of situation and dispositional factors (Endler & Magnusson, 1976; Pervin, 1987; Pervin & Lewis, 1978; Schneider, 1983). Schneider (1983) explicates how many forms of interaction may be described and formally tested. I am concerned with statistical interactions involving continuous variables in the present research investigation.

A statistical interaction reflects a formal test of a hypothesis such that the relationship of one variable to another variable changes based on varying levels of a third variable (Aiken & West, 1991). For example, suppose one were interested in the effect of climate for competitiveness on certain perceptions of individuals, and that it was hypothesized that the effect would vary with level of dispositional competitiveness. More explicitly, individuals who are less competitive become more sensitive to the supportive behavior of coworkers when the climate becomes more competitive. This interaction hypothesis could formally be tested via the regression of perceptions of coworker support behavior on a measure of climate for competitiveness, dispositional competitiveness, and their product. In particular the multiple regression equation to be assessed is:

The model journal for this dissertation is *Psychological Methods*.

$$Y = \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon \,, \tag{1}$$

where Y represents the outcome variable coworker support, X represents the focal variable climate, Z represents the moderator variable disposition competitiveness, XZ represents the product of climate and disposition, the βs represent regression coefficients, and ε is the error in the regression equation. A significant effect for the product term, β_3, would indicate that the effect of climate on perceptions of support is dependent upon[1] the level of dispositional competitiveness. Knowing this, one would want to examine the exact form of the relationship before making recommendations to firms regarding policy decisions such as developing reward structures, incentive plans, and the like. Failure to test for the hypothesized interaction could lead to erroneous conclusions. A zero correlation between two variables does not necessarily indicate there is no relationship between the two variables. The detailed methods for examining interaction (or equivalently, moderation) in the context of multiple regression have been provided by a number of authors (Aguinis, 2004; Aiken & West, 1991; Jaccard & Turrisi, 2003; Jaccard, Turrisi, & Wan, 1990).

In the remainder of this section I will describe statistical difficulties with assessing interactions as well as methods and procedures that have been developed to overcome these difficulties. Finally, I will describe at least two alternatives that should improve upon current methods for detecting interactions: the use of parcels in structural equation modeling and the use of latent variable scores. I will argue that each of these procedures should 1) improve upon the detection of the interaction term, 2) improve upon the model fit within the structural equation modeling framework, and 3) reduce the

[1] Here the terms dependent upon, contingent on, moderated by, or interacts with are used synonymously.

complexity involved in estimating such models. This argument will then lead to two studies involving archival and simulated data. In Study 1, I examine variability in the procedures. In Study 2, I systematically manipulate the use of parcels by varying item-to-parcel ratios, and compare the use of latent variable scores to the method developed by Jöreskog and Yang (1996).

Difficulties in the Detection of Statistical Interactions

Experimental versus observational designs. In demonstrating the effect of a statistical interaction, one may either utilize an experimental or observational design. The difference between the two involves the manipulation of the interaction effect (i.e., experimental design) versus measuring the effect (i.e., observational design). McClelland and Judd (1993) argued persuasively for the use of the experimental design, and in particular a method termed extreme-group designs, because of problems in the measurement of and detection of interactive effects. Cortina and DeShon (1998) countered McClelland and Judd (1993) and maintained that often "in applied psychology … the relative size or importance of an effect" (p. 799) must be estimated for the situation. In short, extreme-group designs artificially inflate effect sizes, whereas the observational method more closely approximates population values. Cortina (2002) argues that experimental designs have more power but at a cost to generalizability. From these arguments, one could conclude that design should be driven by the research question. Mere detection of an effect could be demonstrated by experimental manipulation, but estimation of strength of relationship should come through the observational method (Cortina & DeShon, 1998). However, the use of observational design has its own problems. In particular, measurement errors are present in the predictors and these errors tend to obscure the detection of the interaction effect.

Measurement error and product terms. While adequate psychometric properties are fundamental to any applied research design, the presence of measurement errors greatly exacerbates the problem of the detection of statistical interactions. This problem is manifest in two important ways. First, Borhnstedt and Marwell (1978) demonstrated the influence of the reliability of the component terms on the reliability of the product term. The reliability of a product term is a function of the means, standard deviations, reliabilities, and correlation between the component terms (Borhnstedt & Marwell, 1978). If the correlation between X and Z in Equation 1 is zero and X and Z are mean centered, then the reliability of the product term is the product of the component reliabilities. It can be demonstrated that even when reliabilities of the component terms are relatively high, the product term has a great deal of measurement error (e.g., low reliability). Second, the product reliability directly impacts the increment in the squared multiple correlation (ΔR^2) due to the product term that shows a statistical effect for the interaction (Busemeyer & Jones, 1983). Busemeyer and Jones (1983) demonstrated that the observed effect of the interaction is directly attenuated by the reliability of the product term.

Multiplicative Structural Equation Models

Following the arguments set forth by Borhnstedt and Marwell (1978) and Busemeyer and Jones (1983) as described above, Kenny and Judd (1984) developed a method for assessing multiplicative relationships within a structural equation modeling (SEM) framework. Their model was developed based on the work by Borhnstedt and Goldberger (1969), which detailed the exact relationship of the product variances to the component variances and covariance. Namely, given bivariate normality and scores that are mean centered, three relationships exist: (1) the variance of a product of two random variables must equal the product of the variances of the random variables plus the squared covariance of the random variables; (2) the covariance of the product with either random variable is zero; and (3) the mean of the product term must equal the covariance of the component terms.

Kenny and Judd (1984) demonstrated their method using the simple example of a latent variable *Y* having only one indicator, and two latent variables *X* and *Z* each having two indicators. They argued that the combination of products for each of the indicators of *X* and *Z* can be used to indicate the latent product term (i.e., *XZ*) and that the proper estimation of this model requires nonlinear constraints.

Although the Kenny-Judd method was successful in retrieving the regression coefficients in a simulation, the procedure went largely unused for over a decade (Jaccard & Wan, 1995; Jöreskog & Yang, 1996; Jöreskog, 1998). One reason for this, aside from the statistical complexity of the models, was that the available software was incapable of imposing nonlinear constraints (Jaccard & Wan, 1995). Since the development of LISREL 8, nonlinear constraints are readily implemented, and the 1990s saw a resurgence of interest in the estimation of multiplicative relationships in structural equation modeling (see Jöreskog, 1998 and Schumacker & Marcoulides, 1998 for reviews).

The Kenny-Judd method worked largely because it utilized all possible product combinations of the component indicators to develop the indicators for the latent product term. However, for a structural equation model with several latent variables and several indicators of these variables, the Kenny-Judd method becomes unwieldy. Jaccard and Wan (1995) furthered the Kenny-Judd model in that that they used three indicators each for the component terms instead of two. To reduce model complexity in estimation, they used only two (out of the three possible) indicators of each of the component terms to estimate four product indicators. While Kenny and Judd had only four indicators of the product term, these were a function of all possible combinations of the products of the component indicators. All information from the component terms was utilized by Kenny and Judd, whereas only four out of a possible nine product indicators were used by Jaccard and Wan (1995).

Jöreskog and Yang (1996) built upon the Kenny-Judd model in a number of ways. First, Jöreskog and Yang (1996) argue convincingly that the estimation of such models with nonlinear effects should include mean intercepts. This information is necessary because of additional constraints not previously estimated. Jöreskog and Yang (1996) state that "the means of the observed variables are functions of other parameters in the model and therefore the intercept terms have to be estimated jointly with all the other parameters" (p. 58). By utilizing the means of the indicators, model complexity is greatly increased. Jöreskog and Yang (1996) further demonstrated that only one indicator of the product latent term is necessary for model identification. Herein lays the problem. If the component variables have a large number of indicators, using only one indicator from each component to develop an indicator for the product effectively discards information from all the other component indicators – even if the indicators are the best indicators for each latent component term.

Jöreskog (Jöreskog, 1998; Jöreskog & Yang 1996) has called these methods full-information methods but did not recognize the loss of information from the failure to use the remaining component indicators. Full-information methods estimate all parameters simultaneously (e.g., factor loadings, error variances, structural parameters; Jöreskog, 1998).

Yang (Jöreskog & Yang, 1996; Yang Jonsson, 1997) compared the use of one versus four product indicators. In Jöreskog and Yang (1996) the conclusion was that the difference was minimal, but that the reduction in complexity by using only one indicator outweighed the deficiency incurred by the use of four product indicators. Yang Jonsson (1997) systematically investigated the use of one versus four indicators. The conclusion was that four indicators have less bias in parameter estimation than using one indicator, but that model complexity leads to extremely poor fit and severe underestimation of the standard errors. The four-indicator method probably resulted in better parameter estimation of the product regression coefficient because more information was utilized when indicating the latent product term.

One can demonstrate the likely rationale for the Yang Jonsson (1997) conclusion by taking two latent variables each measured by several items and examining the bivariate correlations among the items. It is unlikely that the items are correlated equally. If one were to choose two indicators such that the product was not sufficiently correlated with the outcome, but the remaining correlations were, then bias would occur in estimation of the coefficient for the product term. The argument that there are many combinations of single indicators has been made elsewhere (Lee, Song, & Poon, 2004).

Following the work of Yang, Jöreskog (1998) has since stated that the full-information methods are quite difficult in practice and require extremely large sample sizes if estimation methods other that maximum likelihood are to be used. Further, Jöreskog states that other procedures are developing (e.g., two-stage least squares, factor scores) that should be improved upon and systematically investigated. Finally, Jöreskog argues that strong theory should guide the use of full-information procedures such as that of Jöreskog and Yang (1996).

The mathematical derivations of the Jöreskog and Yang (1996) procedure have been detailed elsewhere (Jöreskog & Yang, 1996; Yang Jonsson, 1997, 1998; Yang-Wallentin & Jöreskog, 2001). The statistical assumptions and constraints of the model are summarized in Appendix A.

Robust Standard Errors and Chi –Squares. One of the fundamental assumptions for any structural equation modeling analysis that utilizes maximum likelihood estimation is the assumption of multivariate normality (Jöreskog & Sörbom, 1996a; Jöreskog & Yang, 1996). In principle, violating the assumption does not lead to biased parameter estimates but does lead to asymptotically incorrect standard errors (Yang-Wallentin & Jöreskog, 2001). Even if all of the component variables follow a multivariate normal distribution, their products are not distributed normally (Kendall & Stuart, 1958; Kenny & Judd, 1984). For this reason, Jöreskog and Yang (1996) demonstrated the use of three estimation methods: maximum likelihood, weighted least squares, and weighted least squares with an augmented moment matrix. The latter two are supposedly distribution free, meaning that the assumption of multivariate normality is not required. Problems arise in the distribution free methods in that extremely large sample sizes are required (e.g., $N > 3000$) – sample sizes that are unlikely in the behavioral and organizational sciences. Yang Jonsson (1997) investigated these three methods for estimating the Jöreskog-Yang procedure via simulation. She found that maximum likelihood worked fairly well under most sample size conditions (e.g., $400 \leq N \leq 3200$). She and Jöreskog later investigated the effect of using the Sattora-Bentler (1988) correction formula, as a new feature of LISREL (Jöreskog, Sörbom, du Toit, & du Toit, 2000), to correctly estimate the theoretically biased standard errors and chi-square distributions (Yang-Wallentin & Joreskog, 2001). They discovered that the correction did not greatly affect the standard errors or chi-squares for the one-indicator product model. Further, the bias for the one indicator model was smaller than expected.

Similar evidence has been provided elsewhere as to the robustness of maximum likelihood to violations of multivariate normality (Bollen, 1989; Chou, Bentler, & Satorra, 1991; Hu, Bentler & Kano, 1992). Cortina, Chen, and Dunlap (2001) reviewed the literature on the robustness of maximum likelihood to such violations and concluded that structural equation modeling that includes product terms should be relatively robust when estimated with maximum likelihood. Jaccard and Wan (1995, 1996) have made similar conclusions with respect to the inclusion of product terms. Given this evidence, it appears that the Jöreskog-Yang procedure, with one product indicator can safely be estimated via maximum likelihood, but one should always retain some caution and assess the data for severe departures from multivariate normality.

On the issue of centering. It is well known that when estimating regression models involving interaction terms that centering can be useful in reducing the effects of multicollinearity and improving interpretation (Aiken & West, 1991). In fact, Kenny and

Judd (1984) and Jaccard and Wan (1995) indicate that centering is necessary. The Jöreskog-Yang model does not require centering of the indicators to assess the latent variables. The constraints imposed on the intercept terms render the latent variable components mean centered and the latent variable product is likewise constrained to have the relationships that would be obtained by centering. Jöreskog and Yang (1997) use this argument as a further reason for their method. However, centering can still prove useful (Cortina et al., 2001). For instance, unless the component variables are bivariate normal, even with mean centering, the component terms are still correlated with the product term – a constraint imposed in the Jöreskog-Yang procedure (see Appendix A). Jaccard and Wan (1996) adopted the Jöreskog-Yang approach but noted that they were unable to get certain models to converge without mean centering the indicators. In practice, mean centering may reduce many of the constraints to non-significance. While mean centering may be theoretically unnecessary (Jöreskog & Yang, 1996), in practice, centering the indicators prior to creation of product indicators will greatly improve model estimation with the maximum likelihood method.

There are other methods for estimating nonlinear equations (i.e., multiplicative models) that attempt to reduce the affects of measurement error. These methods have been reviewed in Schumacker and Marcoulides (1998) and by Cortina et al. (2001). Among these various procedures are methods involving two-stage least squares estimation (Bollen, 1995, 1996; Bollen & Paxton, 1998) and two-step procedures (e.g., Mathieu, Tannenbaum, & Salas, 1992; Ping, 1995, 1996). All of these procedures are based in whole or in part on the Kenny-Judd model. For the purposes of this dissertation, however, I focus and build on the work by Jöreskog and Yang (Jöreskog, 1998; Jöreskog & Yang, 1996; Yang Jonsson, 1997; Yang-Wallentin & Jöreskog, 2001).

Item Parcels as a Potential Solution

The measurement of constructs in the behavioral and organizational sciences is often accomplished by the use of a rating scale composed of several items (Bagozzi & Edwards, 1998). There exist a number of ways to represent the latent variables measured by these items. One such method is aggregating items into parcels prior to analysis. "A parcel can be defined as an aggregate-level indicator comprised of the sum (or average) of two or more items, responses, or behaviors" (Little, Cunningham, Shahar, & Widaman, 2002, p. 152). The use of item parcels in factor analytic methods such as structural equation modeling has a long and tumultuous history. Indeed, two recent reviews (Bandolos & Finney, 2001; Little et al., 2002) discuss the controversial nature of the use of parcels. Little et al. (2002) describe the controversy as follows:

> To some, aggregating items to manufacture indicators of constructs is viewed as a dubious practice at best and cheating at its worst. Moreover, the practice of parceling contributes to the oft-whispered reputation of SEM [structural equation modeling] as yielding a "smoke-and-mirrors" distortion of reality. For advocates of parceling, on the other hand, the practice is viewed as one that puts a fine sheen on an otherwise cloudy and therefore difficult to discern picture of reality. In this sense, the use of parcels in SEM is not seen as invoking smoke and mirrors, but rather as providing a carefully polished mirror of reality that really "smokes" (p. 152).

Dimensionality as a source of conflict. With few exceptions, most methodologists would agree that a set of items to be parceled should be unidimensional and relatively free from unwanted sources of shared variance (cf. Bagozzi & Edwards, 1998; Bandalos & Finney, 2001; Kishton & Widaman, 1994). A brief review of classical test theory will assist in illustrating this point. Any observed score for item *i*, can be conceptualized as having the following components:

$$X_i = T_i + S_i + e_i, \tag{2}$$

where X is the observed score, T is the true score for the latent variable, S is a source of systematic variance unrelated to the latent variable of interest, and e represents random measurement error. The goal of factor analysis is to partition the observed variance in a set of items into their common and unique sources of variance. The common source is that due to the latent variable, whereas the unique source is that due to systematic and error variance. When structural equation modeling is not used, the items are typically averaged to represent the latent variable. This assumes a unit weighting for each item and the observed score for the latent variable contains elements of the true score plus any systematic and unique error. This principle applies whether subsets of items or all items are averaged to represent the construct.

Hall, Snell, and Singer-Foust (1999) demonstrated that small, modest sources of shared systematic variance had dramatic influences on parameter estimation in structural equation modeling. Bandalos (2002) demonstrated that shared sources of variance unrelated to the latent variable of interest influence parameter estimation as well as fit indices. If the unique sources of variance are shared across parcels, then the variance is not removed when estimating the latent variable. The shared variance across parcels is subsumed into latent variable variance. The subsumed variance defines the *distributed uniqueness strategy* for forming parcels (Hall et al., 1999). In contrast, Hall et al. (1999) recommended an *isolated uniqueness strategy* for the formulation of parcels by placing items that share unique variance into the same parcel. Any shared systematic variance can then be separated from the latent variable variance (i.e., it is represented in some but not all parcels).

A problem with the isolated uniqueness strategy of Hall et al. (1999) is in determining the presence of an unrelated latent variable (e.g., social desirability). Hagtvet and Nasser (2004) used a second-order factor analysis strategy to compare isolated versus distributed uniqueness strategies. Their procedure involves systematically examining the modification indices for the error variance in a confirmatory factor analysis where the latent variable is the second-order factor, the parcels are the first-order factors, and the items are the manifest indicators. This approach seems reasonable. However, the work by Bandalos (2002), Hall et al. (1999), and Hagtvet and Nasser (2004) assume the presence of a secondary nuisance factor and that its influence is not shared by every indicator as might be the case for a method factor.

Benefits of using parcels. In reviewing the merits of using parcels, Little et al. (2002) identified two broad categories: psychometric and model-level considerations. Psychometric considerations concern the item or indicator properties relative to the latent construct of interest. Items tend to be less reliable than aggregate indicators. Further, items drawn from a more diverse domain will be less efficient and therefore have lower communalities than will aggregate indicators (Little, Lindenberger, & Nesserlroade, 1999). This conclusion assumes that the aggregate indicators are not unduly influenced by nuisance factors in what Little et al. (2002) describe as "dirty measures." As described above, the potential of shared systematic variance has lead many researchers to be less than optimistic about using aggregate indicators.

Items also have a greater likelihood of distributional violations compared to parcels. Aggregate indicators are more likely to be distributed normally. Items have fewer, larger, and less equal intervals between scale points than do parcels. An item measured on a 5-point scale has only five possible values, whereas an average (or sum) of three such items has 13 possible values. Therefore, parcels will resemble continuous variables more so than will items.

Using parcels has several benefits for evaluating structural equation models. While there are several benefits for the use of parcels, I will focus on a few of these benefits as they might apply to the Jöreskog-Yang method for estimating latent variable interactions. Models based on parcels as compared to single items are more parsimonious in that there are fewer parameters to estimate. For instance, three latent constructs each measured with six items renders a covariance matrix to be analyzed with $(18*(18 + 1))/2 = 171$ entries. The same model with parceled data (e.g., three parcels each with two items) has $(9*(9 + 1))/2 = 45$ entries. Models based on parceled data also have fewer chances for residuals to be correlated or for dual loadings to emerge (Little et al., 2002). Analysis of item level data in comparison to parceled data has been shown to require more iterations to converge and have larger standard errors leading to poor fitting models (Little et al., 2002). Other simulations have refuted this last benefit (Marsh, Hau, Balla, & Grayson, 1998; Nasser & Wisenbaker, 2003). For example, Nasser and Wisenbaker (2003) found that models based on fewer parcels (i.e., more items per parcel) had better fit indices, but these same models had more incidences of nonconvergence at smaller sample sizes. Marsh et al. (1998) found that constructs were more accurately measured with more indicators than fewer. However, Little et al. (2002) question the generalizability of the Marsh et al. (1998) findings due to the narrowness of the simulation.

A problem with assessing latent variable interactions is model complexity. The sheer number of statistical constraints that can be required leads to a high potential for model misfit (e.g., slight violations of multivariate normality have large impacts on fit measures). The greater the number of indicators, the greater the number of constraints, and the more likely model misfitting can occur. Although the Jöreskog-Yang approach of using a single indicator to represent the latent product addresses the effects of model complexity, the reliance on single items raises concerns of the adequacy of measuring the latent variables. It is hypothesized that, by carefully parceling items one can improve the likelihood of accurately assessing the product via a single parcel and simultaneously removing the effects of measurement error known to attenuate the product effect.

Methods for parceling items. Both of the recent reviews of parceling research (Bandalos & Finney, 2001; Little et al., 2002) provide descriptions of methods for parceling items. The methods described do not completely overlap. The rationale for parceling data should be clear and explicit (Bandalos & Finney, 2001; Hall et al., 1999). Further, the reason for parceling should drive the choice of the procedure to use. For instance, if a researcher's goal is simply to improve the normality of indicators, then items should be placed together that will "cancel out" the skewness of other items. Some researchers advocate parceling based on item content (Comrey, 1970 as cited in Bandalos & Finney, 2001). For the purposes of this discussion, I will focus on empirical methods that should be useful in estimating latent variable interactions.

Cattell (1974; Cattell & Bursdal, 1975) advocated a method he called *radial parceling*. The two-step procedure involved a factor analysis of the items, and then forming parcels based on congruence coefficients of the factor loadings. The procedure was designed to be useful in factor analytic studies such as those influential in determining factor structures of personality. One problem with this approach, aside from being labor intensive, is the possibility of items that reflect different factors (i.e., latent variables) being placed together in the same parcel (see Bandalos & Finney, 2001; Barrett & Kline, 1981). This radical parceling procedure is most useful in complex multifactor datasets.

Kishton and Widaman (1994) describe a *random* procedure. Little (Little et al., 1999; Little et al., 2002) has argued that randomly assigning items to parcels may be better than using items themselves under certain circumstances. Analyzing parcels by random assignment presumes that parceling strategy makes no difference. However, parceling strategy does have an influence on the measurement of the latent variable. Random assignment may not lead to the most accurate assessment of the latent variable.

Little et al. (2002) describe a method that may be useful for developing parcels that are nearly parallel as indicators of the latent variable. The *item-to-construct balance* approach involves alternating assigning items to parcels based on item factor loadings. For example, to form 3 parcels of 2 items each, one takes the three highest loading items and assigns them the first three parcels, respectively. Then, one takes the three lowest loading items and assigns them in reverse order to the three parcels, respectively. Thus, the highest and lowest loading items form the first parcel, the next highest and next lowest loading items form the second parcel, etc. The *item-to-construct balance* approach should be used if and only if parallel indicators are desired. With the approach, the negative influence of measurement error is distributed equally among the parcels thereby making structural equation modeling ineffective at partialing out unique error variance. The result is a smoothed out analysis, but at the risk of distributing uniqueness across parcels similar to that done by unit weighting the items.

The final method I will describe, to my knowledge, has not been described in the literature on creating parcels. Rather than assume the presence of a nuisance factor, or the need to balance the indicators, one can create parcels that isolate the most similar items in terms of their relationship to the latent factor. The procedure involves placing the items with the most similar standardized factor loadings into the same parcels. For example, if six items are to be parceled into three indicators, the highest two loading items will go into parcel 1, the next two highest will go into parcel 2, and the lowest two items will go into parcel 3. Several assumptions must be made for this procedure to be viable. First, the items must be unidimensional to ensure that the items will correlate reasonably well within each parcel. Second, reasonable care should be given that the items within a parcel are homogenous (Kishton & Widaman, 1994; Little et al., 2002). It should be noted that the standardized loading to be used for parceling the items are from a completely standardized solution of a confirmatory factor analysis. That is, they are the square root of the squared multiple correlation for the item. The squared multiple correlation includes the factor loading and the error variance (i.e., SMC = factor loading^2 / [factor loading^2 + error variance]). The procedure is similar to Catell's radial parceling procedure as well as Hall et al.'s (1999) isolated uniqueness procedure. However, this method presumes neither a multi-factor set of items nor the presence of some nuisance factor. The items will be reasonably homogenous within each parcel, but congeneric across each parcel. The marker indicator should be the one that contains the highest loading items and therefore the least amount of unique variance. The marker indicators are the ones used to form the single product indicator for the Jöreskog-Yang method. In preliminary simulations, this *congeneric method* to parceling items more closely approximates item level estimation of the latent factor, whereas the item-to-construct balance approach to parceling items more closely approximates a unit-weighting of the items.

For the Jöreskog-Yang method, the choice of which indicators to use for formation of the product indicator should be of little importance if parcels are formed using the item-to-construct balance approach. The parcels are all reasonably parallel in their relationship to the latent factor. However, the balanced approach may not remove the influence of measurement error and therefore may not reduce bias in the estimation of the product coefficient (i.e., γ_3 in LISREL terminology – see Appendix A).

The congeneric indicator approach should 1) remove the ill effects of measurement error by isolating the best items for the measurement of the latent factors, and 2) combine information from multiple items to increase the likelihood of efficiently measuring the latent product term leading to less bias in the product coefficient. This effect should be greatest for scales with large numbers of items. For example, twelve items aggregated into three parcels uses the four 'best' items for indicating the latent product, simultaneously parceling out effects of error.

In sum, the present research investigates two strategies for forming item parcels. The strategies are the item-to-construct balanced approach and the congeneric approach. For each of these strategies, item-to-parcel ratios will also be assessed. These strategies are compared for their effectiveness in the Jöreskog-Yang method of detecting latent variable interactions.

Latent Variable Scores as a Potential Solution

The estimation of factor scores via exploratory factor analysis has a long history (e.g., Lawley & Maxwell, 1971). As a new feature of LISREL, Jöreskog (2000) has developed a procedure for estimating factors scores that he calls latent variable scores. In this section, I describe its role in estimating latent variable interaction effects.

Yang Jonsson (1998) utilized factor scores to create a factor product term and subsequently analyze the relationship treating the factors scores as observed variables. The procedure was thought to remove measurement error from the regression for detecting interactions, but the procedure has known flaws (Bollen, 1989; Yang Jonsson, 1998). Namely, factor score estimates do not represent the factors precisely (Bollen, 1989). Most factor score procedures can render scores that account for little variance in the factors themselves. Nonetheless the procedure is worth pursuing (Jöreskog, 1998). The factor scores procedure was helpful in illustrating the simplification of the estimation of nonlinear equations and led Jöreskog (1998) to call for a systematic investigation of the use of factor scores in estimating non-linear equations.

As of LISREL 8.3, the program has been able to compute latent variable scores (Jöreskog et al., 2000). Latent variable scores (LVS) can be distinguished from factor scores in that LVS will "satisfy the relationship of the latent variables themselves" (Jöreskog, 2000, p. 1). The mathematical derivations of the LVS are provided by Jöreskog (2000). One can verify that the results obtained by first developing LVS and then treating them as observed variables is the exact same result as if one were utilizing a full-information model via LISREL (i.e., structural equation model). The structural parameters will be identical; however, Jöreskog (2000) cautions that the standard errors may not be identical.

The use of latent variable scores is a promising procedure for a number of reasons. First, it explicitly separates the measurement model from the structural model, a practice that has been called for elsewhere (McDonald & Ho, 2002). Second, one can investigate the distribution of residuals of structural equations that could not be done via LISREL prior to this implementation. Third, one can create nonlinear functions of latent variables (e.g., squares, products) without implementing complex constraints. It is this third feature that is most relevant here.

Jöreskog (2000; Jöreskog et al., 2000) has described how to compute latent variable scores via LISREL and has provided examples of their use. Because the procedure is not a common practice, I will briefly reiterate how they are computed and why they may be useful in improving the detection of interactions in structural equation modeling. The first step is to conduct a confirmatory factor analysis on all the relevant variables in the model system (e.g., the indicators for Y, X, and Z in Equation 1, see pg. 2). From this analysis, LISREL appends the latent variable scores to the system file containing the raw data for the indicators. The nonlinear function is then computed from the latent variable scores (e.g., the latent variable scores for X are multiplied by those for Z to create the product variable, XZ). These latent variable scores are now treated as observed variables.

In the latent variable scores procedure, measurement error is removed from the component variables prior to the creation of the product, whereas in full-information methods, the indicators for the products themselves contained error. In full-information methods there are no 'indicators' for the latent product. Further, the complication of how the error terms for each indicator relate to the latent product is also removed. In contrast to the Jaccard-Wan and Jöreskog-Yang methods, all indicators for the component latent variables are utilized in determining the latent product when using latent variable scores. The efficiency and the extent to which this procedure leads to biased estimates are not yet known. Given the use of information from all the indicators, bias should be minimal, but as Jöreskog (2000) mentions, the standard errors may not be the same as using full-information methods. The relative efficiency of this method is investigated in Study 2.

Purpose of the Research

The purpose of this research is to investigate the effects of the use of parcels in structural equation modeling and the use of latent variable scores in detecting interactions. I have argued that each of these procedures should 1) improve upon the measurement of the interaction term, 2) improve upon the structural equation model fit, and 3) reduce the complexity involved in estimating models involving interactions. Two studies are designed that involve real and simulated data. The purpose of Study 1 is to examine the variability in the procedures in detecting a latent variable interaction using real data. In Study 2, I systematically manipulate the use of parcels by varying item-to-parcel ratios, and compare the use of latent variable scores to the method developed by Jöreskog and Yang (1996).

STUDY 1

AN EMPIRICAL EXAMPLE

Background

Over the last 15 to 20 years organizations have been moving towards team-based structures which require more collaborative strategies for work (Kozlowski & Bell, 2003). The phenomenon is of obvious importance to the organizational sciences. However, much of the research on the effects of collaborative (versus competitive) strategies has been in social or educational psychology and not in the organizational sciences. Consequently, current practice in organizations doesn't match known theory and research. In fact, a change from competitive to collaborative strategies of work may not be an easy sell to managers of organizations.

One rationale for this resistance to change is that the culture in the U.S. predominately advocates a competitive style of working despite the move toward more collaborative organizational structures. Another rationale is the seemingly mixed effects of competitive practices on performance. Much of the research in the academic community is definitive (see meta-analyses by Johnson & Johnson, 1989 and Stanne, Johnson, & Johnson, 1999). Social support suffers under competitive arrangements. What have been neglected in this body of research are the effects of a competitive climate as moderated by trait competitiveness.

It is plausible to hypothesize that individuals who are less competitive may be more sensitive to others' behavior in competitive environments. Less competitive individuals may notice the lowered support from coworkers in highly competitive environments. Highly competitive individuals on the other hand may be less prone to notice the effects of competitive practices (e.g., reward distributions based on relative performance). Highly competitive individuals may not perceive differences in the level of support received from their coworkers due to environmental changes in competitiveness. Indeed, competitive individuals may be more focused on their own behavior rather than that of others. The hypothesis that a competitive climate may reduce perceptions of levels of support received from coworkers and that the perception of this reduction in support is contingent upon a person's level of trait competitiveness. This hypothesis requires a test of a statistical interaction.

The purpose of Study 1 is to assess the hypothesis using various methods including: the Jöreskog-Yang approach, and the use of latent variable scores. Three methods of item parceling will be compared (i.e., item level, item-to-construct balance, congeneric parceling). Ultimately, the study will show that there is indeed variability in the procedures and that a researcher is left wondering which results to rely upon. If consistent results are shown, then any approach could be reported. The most parsimonious would be preferred. The particular variables in this study were chosen (1) out of theoretical interest to the author, and (2) because it is thought that they have a sufficiently small effect size to warrant the methods proposed here. That is, if variables were used such that the effect size was large, variability in the procedures may go undetected. I believe that the difficulties encountered in Study 1 (e.g., real data with small effect sizes) are representative of research in the behavioral and organizational sciences.

Method

Data. Two types of coworker support (instrumental and affective), each consisting of five items, serve as dependent variables in separate sets of analyses

(Ducharme & Martin, 2000). The independent variables are trait competitiveness (Helmreich & Spence, 1978) and a modified version of the competitive psychological climate scale used by Brown, Cron, and Slocum (1998), each consisting of four items. The competition items were on a response scale ranging from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*). Coefficient alpha for the scales are .82 and .88, respectively. The coworker support items were on a response scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). Coefficient alpha for the scales are .92 for instrumental and .92 for affective support. The items were mean centered prior to analysis. The items are presented in Appendix B. These data were collected as part of a large-scale project aimed at understanding the relative under representation of women and minorities in the information technology (IT) workforce (Major et al., 2003). In all, 916 information technology workers from 11 organizations responded to a web-based survey. There are 863 complete responses to the 18 items.

Hypothesis tests. The basic model to be evaluated involves the test of the statistical interaction of trait competitiveness and competitive psychological climate in influencing perceptions of support from coworkers (instrumental and affective). The hypotheses can be represented as follows:

Model 1: Instrumental support = trait + climate + trait×climate

Model 2: Affective support = trait + climate + trait×climate

The hypothesized statistical test of interest is the regression coefficient for the interaction term (*trait×climate*) in both models.

Design. To test each of the hypotheses, six methods of estimation were employed. The Jöreskog-Yang method with one product indicator was compared to the use of latent variable scores. Within each of the above procedures, various parceling strategies were employed: item level (no parcels), item-to-construct balance parcels, and congeneric parcels. A sample program for testing this model via the Jöreskog-Yang method is presented in Appendix C.

Outcomes examined. For each of the six analyses on instrumental and affective support, there were both model level and individual parameter level outcomes to consider. At the model level, various fit indices were assessed for the confirmatory factor analyses for the latent variable scores, and for the Jöreskog-Yang method. First, the model ratio of χ^2 to degrees of freedom was compared to its expected value of 1. In addition to the χ^2 tests, the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), the Akaike information criterion (AIC; Akaike, 1987), the non-normed fit index (NNFI; Tucker & Lewis, 1973), and the comparative fit index (CFI; Bentler, 1990) were also assessed. Various criteria for the evaluation of these fit indices have been provided over the years. Hu and Bentler (1999) recently evaluated these criteria and provided some recommendations for cut-off criteria. However, the concern in the present study is whether one procedure is starkly different from the others in terms of model fit. I am interested in relative comparisons rather than absolute cut-off criteria. Because the models differ for the latent variable scores (i.e., only a measurement model) and Jöreskog-Yang (e.g., complex structural and measurement models) methods, the model level information is of only modest importance. The major concern involves the differences between the parceling strategies within each of the estimation procedures.

In addition, the structural coefficient for the product term was examined. That is, from Equation 1 (p. 2), I am interested in the significance of β_3. In LISREL terminology (see Equation A.1, p. 79), the coefficient is labeled γ_3. Ultimately, I am interested in whether the methods produce similar results leading to the same conclusions. The coefficient for the interaction term was examined in terms of statistical significance (i.e., *t* test) and effect size (f^2). The *t* test and f^2 are functionally related to one another, but they are reported because of their familiarity to multiple regression researchers.

Results

Descriptive statistics are presented in Table 1. The coworker support variables are similar with respect to means and standard deviations, and they are highly correlated, r = .73. The support variables are not correlated with the competition variables equally. They are treated separately because the magnitude of the influence of the interaction of the competition variables with each support variable differs (see below).

Table 1
Means, Standard Deviations, and Correlations Among Study 1 Variables

		Mean	SD	1	2	3
1	Affective Support	3.86	0.74			
2	Instrumental Support	4.03	0.73	.73		
3	Competitive Psychological Climate	3.84	1.27	-.08	-.09	
4	Trait Competitiveness	4.43	1.39	.06	-.02	.28

Note. N=863. Correlations above |.08| are significant, $p < .05$.

The model fit indices are displayed in Table 2 for each of the study conditions. For all conditions, the fit statistics would suggest that the interaction model shows a good fit to the data. For the Jöreskog-Yang method, the item level analyses (i.e., use of no parcels) have slightly poorer fits than either parcel strategy (i.e., balanced or congeneric). The fit statistics from the confirmatory factor analysis results used to generate the latent variable scores are also displayed in Table 3. A general trend emerges for the parceling strategies across each outcome variable. The balanced approach has a lower χ^2/df ratio, lower RMSEA, and lower AIC than the congeneric approach indicating a slightly better fitting model.

Table 2
Model Fit Indices for Study 1 Analyses

Outcome Variable	Condition	χ^2/df	RMSEA	AIC	NNFI	CFI
Affective	JY-item	3.30	0.05	351.45	0.97	0.98
Support	JY-Bal	1.54	0.02	79.78	0.99	1.00
	JY-Con	1.86	0.03	84.36	0.99	0.99
	LVS-item[a]	**	**	**	**	**
	LVS-Bal	1.88	0.03	41.23	0.99	1.00
	LVS-Con	1.74	0.03	40.60	0.99	1.00
Instrumental	JY-item	3.77	0.06	386.19	0.97	0.97
Support	JY-Bal	1.11	0.01	74.29	1.00	1.00
	JY-Con	2.67	0.04	94.59	0.98	0.99
	LVS-item[a]	**	**	**	**	**
	LVS-Bal	0.76	0.00	34.55	1.00	1.00
	LVS-Con	3.15	0.05	49.02	.98	.99
CFA all 4 Constructs	Item[a]	4.19	0.06	662.10	.97	.98

Note. JY = Jöreskog-Yang method with one product indicator. LVS = latent variable scores method. Item = parcel strategy is item level. Bal = balanced parcel strategy. Con = congeneric parcel strategy. RMSEA = root mean square error of approximation. AIC = Akaike's information criterion. NNFI = non-normed fit index. CFI = comparative fit index.
[a] Fit indices for LVS method refer to the confirmatory factor analysis prior to creation of the product term. For the item level, the latent variable scores were created via the confirmatory factor analysis of all four constructs.

Parameter estimates and effects size indices for the interaction term are presented in Table 3. The effect size is much larger for the interactive relationship on affective support (mean f^2 = .014) than instrumental support (mean f^2 = .004) across the different analytic methods. For affective support, the interaction is significant five out of six times. Only when the Jöreskog-Yang method is coupled with the balanced parceling strategy is the interaction not significant, $p < .05$. For instrumental support, the interaction is significant three out of six times. Again, the balanced parceling strategy appears to have attenuated the interaction effect for both the Jöreskog-Yang method and the latent variable scores method. With the Jöreskog-Yang method, only by using the congeneric parceling strategy did the interaction term appear significant, $p < .05$. Overall, the patterns of the form of the interactive relationship of climate and personality on affective and instrumental support are quite similar across the analytic methods despite the differences in magnitude. The forms of the relationships as estimated by the different analytic methods (and parceling strategies) are depicted in Figures 1 and 2. As shown in these figures, the general relationship is that (a) there is an interaction among trait competitiveness and climate in affecting coworker support, (b) the crossover point for different levels of trait competitiveness is approximately one standard deviation below the mean for affective support and approximately .5 standard deviations below the mean for instrumental support, (c) the magnitude of the interaction effect is greater for affective than instrumental support, and (d) the strength of the relationship of climate to coworker support is relatively weak at higher trait competitiveness and relatively strong at lower trait competitiveness.

Table 3

Parameter Estimates and Effect Size for Interaction Effect in Study 1

Outcome Variable	Condition	B_{INT}	SE	t	f^2	Model R^2
Affective	JY-item	0.078*	0.024	3.28	0.013	0.05
Support	JY-Bal	0.032	0.019	1.67	0.003	0.02
	JY-Con	0.094*	0.025	3.74	0.016	0.07
	LVS-item	0.063*	0.017	3.79	0.017	0.04
	LVS-Bal	0.052*	0.014	3.76	0.016	0.03
	LVS-Con	0.055*	0.014	4.03	0.019	0.04
Instrumental	JY-item	0.038	0.022	1.73	0.003	0.03
Support	JY-Bal	0.002	0.019	0.09	0.000	0.01
	JY-Con	0.054*	0.021	2.59	0.008	0.03
	LVS-item	0.032*	0.014	2.28	0.006	0.02
	LVS-Bal	0.026	0.014	1.93	0.004	0.01
	LVS-Con	0.033*	0.015	2.16	0.005	0.02

Note. JY = Jöreskog-Yang method with one product indicator. LVS = latent variable scores method. Item = parcel strategy is item level. Bal = balanced parcel strategy. Con = congeneric parcel strategy. * $p < .05$.

To summarize the results for Study 1, slight but noticeable variability exists in the various methods with respect to model fit and parameter estimation. Parceling strategy has the following effect on model fit in ascending order of goodness-of-fit, item < congeneric < balanced. In general, all methods produce models with acceptable fit. Parceling strategy has the follow effect on parameter estimation in ascending order of goodness-of-fit, item < balanced < congeneric. The balanced approach appears to produce better fitting models than the congeneric approach but at the expense of attenuated structural coefficients. The use of latent variable scores appears to produce larger effect sizes than the Jöreskog-Yang method at smaller effect sizes (i.e., for instrumental support) as well as smaller effect sizes at larger effect sizes (i.e., affective support) indicating a potential interaction with respect to estimation method and effect size. With only two examples, affective support and instrumental support, the extent to which these findings are generalizable is not known. However, the existence of such variability lends strong support for the need for a simulation study to investigate these relationships.

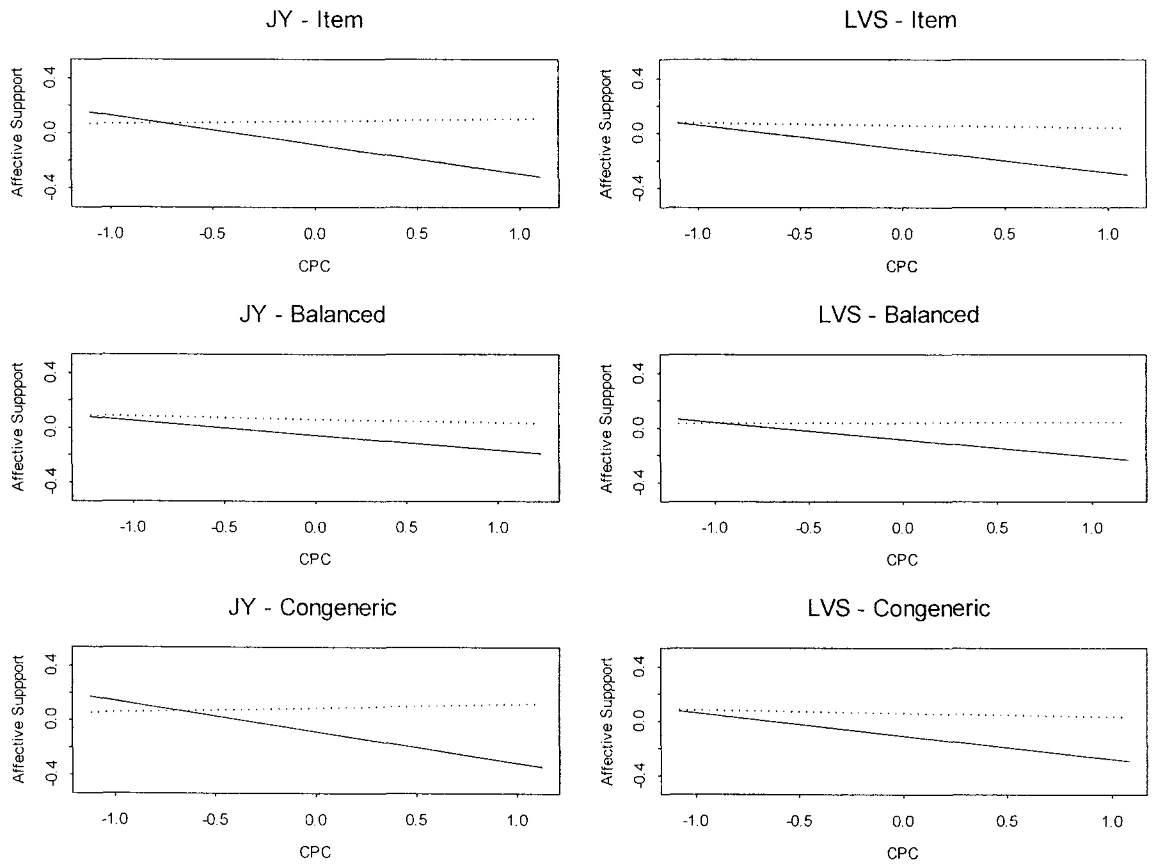

Figure 1. Plot of the interaction for two-levels of trait competitiveness (1SD above the mean – dotted line and 1SD below the mean – solid line) depicting the relationship of competitive psychological climate to affective support for each of the conditions in Study 1.

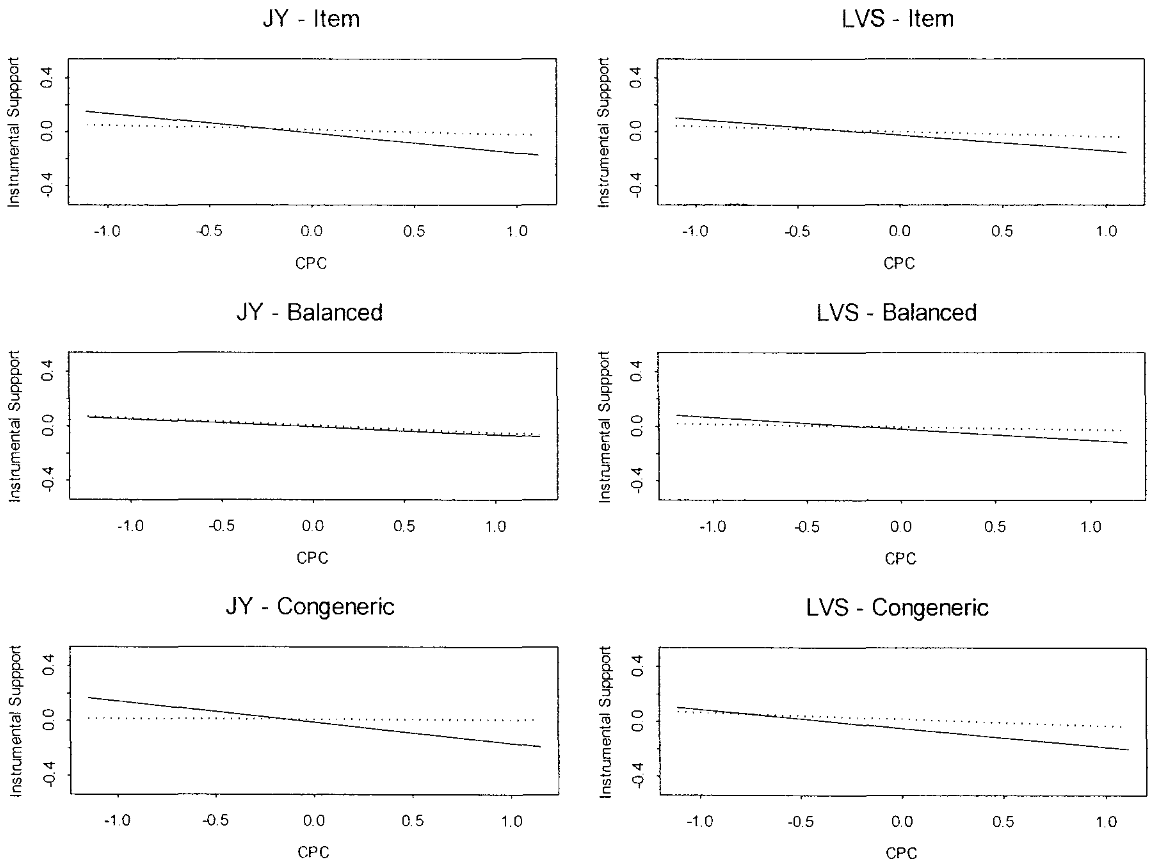

Figure 2. Plot of the interaction for two-levels of trait competitiveness (1SD above the mean – dotted line and 1SD below the mean – solid line) depicting the relationship of competitive psychological climate to instrumental support for each of the conditions in Study 1.

STUDY 2

A MONTE CARLO SIMULATION

Background

While Study 1 may demonstrate variability in the procedures, it does not address which of the procedures produce accurate estimates of the interaction effect. That is, given this procedural variability (i.e., mixed results) and unknown population values, one cannot know with certainty whether a method that rejected the null (i.e., a significant product term) is committing a Type I error or if the effect is present in the population. Likewise, one cannot know with certainty if a method that failed to reject the null is committing a Type II error, or the effect is not present in the population. The effect cannot be simultaneously present and not present in the population. Any discrepancy is therefore due to method of estimation. The purpose of Study 2 is to simulate data with known population values (e.g., interaction effect) and then compare the procedures. The data generated in Study 2 will adhere to normal theory with the exception of the product term. I describe the design below.

Method

Procedure. Data were simulated with PRELIS 2.54 as prescribed by Jöreskog and Sörbom (1996b). Appendix D contains a sample program for generating the data with PRELIS. Each sample generated, consisted of 800 observations. This number was chosen based on 1) it is similar to the sample size (N = 863) in Study 1 and therefore is comparable for the given parameter estimates, and 2) because N = 800 should have sufficient power to detect a small effect size, f^2 = .01 (Cohen, 1988).

Three effect sizes for the interaction effect were examined while holding the $R^2_{(\text{additive})}$ for the model constant at .05: f^2 = .01, .02, and .15. These reflect typical, small, and medium effect sizes in the interaction and social science literature (see Champoux & Peters, 1987; Chaplin, 1991; Cohen, 1962, 1988, 1992).

For each design condition, 500 replications were conducted. That is, 500 samples were generated from the population (Mooney, 1997). There exists no sound argument for the precise number of replications to have in a simulation study. Some have argued as many as 100,000 replications should be generated; others have argued as few as 100 (Mooney, 1997). In structural equation modeling research, most replications have been between 100 and 400 (Yang Jonsson, 1997). Yang Jonsson (1997) chose 600 as an arbitrary number in her doctoral dissertation, presumably to be slightly larger than 400. If a researcher is investigating the tails of a distribution, then larger numbers of replications are needed; if the shape of a distribution (i.e., normality) and central tendency are all that are investigated, then fewer replications are needed (e.g., 100). My choice of 500 seems both logical and reasonable given the purpose of Study 2.

For the latent variables X and Z, 12 items were generated as indicators with the measurement properties shown in Table 4. The items were generated in descending order of squared multiple correlation with the latent variable in order to contrast the parceling strategies. The latent independent variables, X and Z, have the same properties but different random normal variables were used for their generation. Parcels were generated according to the design. To keep the latent dependent variable Y consistent across manipulations, three indicators were generated based on a composite reliability of .90. These three indicators were used for Y regardless of the design cell for each replication.

Design. A total of 1500 samples were generated – 500 for each effect size, f^2 = .01, .02, and .15. Parcels were formed such that the number of parcels (and items per parcel) are: 3(4), 4(3), and 12(1). The same parcel size and strategy were used for *X* and *Z* in each condition; there was no mixing. Both the item-to-construct balance approach and the congeneric parceling approach were contrasted. For each of the 1500 samples of size 800. the Jöreskog-Yang approach with one product indicator was contrasted with the use of latent variable scores. The 1500 samples (3 effect sizes × 500 replications) × 5 parcel conditions × 2 methods requires 15,000 analyses.

Table 4
Measurement Properties of Simulated Population Values

Item	Lambda X	Theta Delta	SMC	SC LX
1	.950	.750	.55	.74
2	.925	.875	.49	.70
3	.900	1.000	.45	.67
4	.875	1.125	.40	.64
5	.850	1.250	.37	.61
6	.825	1.375	.33	.58
7	.800	1.500	.30	.55
8	.775	1.625	.27	.52
9	.750	1.750	.24	.49
10	.725	1.875	.22	.47
11	.700	2.000	.20	.44
12	.675	2.125	.18	.42

Note. Composite reliability for the population values is .85. Lambda X is the latent variable loading for the item. Theta Delta is unique variance for the item. SMC is the squared multiple correlation. SC LX is the completely standardized factor loading.

Outcomes to examine. For the 15,000 analyses, there are both model and individual parameter considerations. For the latent variable scores method, model considerations are only relevant for the confirmatory factor analysis of the individual items (i.e., Y1, Y2, Y3, X1, X2, X3, X4, Z1, Z2, Z3, Z4). No product information was considered at the model level for the latent variable scores method. For the Jöreskog-

Yang approach, there is a possibility of non-convergence in the models (number of iterations is set at 500), and a possibility of admissibility concerns. With respect to admissibility, negative values could be returned for variance estimates due to the maximum likelihood iteration algorithm. The number of non-converged models and non-admissible models was counted. Model fit was assessed. Fit indices examined include: χ^2/df, RMSEA, AIC, NNFI, and CFI. These indices have the same interpretations as described in Study 1.

As in Study 1, parameter estimates for the product term coefficient were compared across conditions. The estimate (γ_3), and effect size (f^2) were examined. The estimates (γ_3) should be normally distributed for the 500 replications. This was evaluated by examining density curves (i.e., smoothed histograms), and tests of skewness and kurtosis. Bias in the estimates is indicated by differences in the population parameter as specified by the research design, and the observed statistic recovered by the method of estimation (Mooney, 1997). Suppose that the parameter to be examined is θ, and the ith observed estimate is $\hat{\theta}_i$, then the bias statistic is computed as $E(\theta)-\theta$ where $E(\theta)$ is the mean of $\hat{\theta}_i$ for all i. The significance of the bias statistic can be assessed statistically (i.e., bias = 0). The ratio of the bias statistic to its standard error is distributed as t (or standard normal with large numbers). The standard error of the bias statistic is computed as:

$$\frac{SD(\hat{\theta})}{\sqrt{R}} \tag{3}$$

Here, $SD(\hat{\theta})$ is the standard deviation in the observed estimates, and R is the number of replications. Another measure of bias commonly assessed is the root mean square error (RMSE). With the variables defined as above, RMSE is computed as:

$$\sqrt{\frac{\sum_{i=1}^{R}(\hat{\theta}_i - \theta)^2}{R}} \tag{4}$$

With respect to statistical inference, the efficiency of the estimation procedures was also examined (Mooney, 1997). Efficiency refers to the variability of a sample statistic. Highly efficient methods will render parameter estimates drawn at random from the same population with low variance. Less efficient methods will have higher variability. Given 500 replications, efficient methods will render 500 parameter estimates with little variability about the mean of the parameter estimates. The empirical standard error in a simulation study is the standard deviation of the observed parameter estimates. That is, the standard deviation of $\hat{\theta}$. One can investigate the bias in the estimation of the standard errors by the standard error ratio: the average of the observed standard errors to the empirical standard error. Ratios above one are indicative of overestimation of the standard errors. Ratios below one are indicative of underestimation of the standard errors (Jöreskog et al., 2000; Yang Jonsson, 1997). Bias in the standard errors has direct implications on the validity of tests of significance for the parameter estimate.

Results

The results are presented in two sections addressing model level outcomes and parameter estimation. The results of the 15,000 simulations are presented in tabular and graphical forms. The figures are presented to provide maximum information in minimal space (i.e., high data-to-ink ratio; Cleveland, 1993; Tufte, 2001). In most instances, it is the shape of the figures (e.g., density curve) that is of interest, not the numerical values themselves.

Model fit. The results of model fit are separated into the results for the confirmatory factor analysis used to create the latent variable scores and the results for the model fit of the Jöreskog-Yang method. Across all simulations, there are no instances of non-convergence or non-admissibility. Table 5 presents the averages for all sets of analyses for each of the outcome variables studied: χ^2/df, RMSEA, AIC, NNFI, and CFI (as described above). The models appear to fit the data well across all study conditions. That is, based on the fit criteria, a researcher is likely to accept the model as a good fit to the data in all study conditions.

While there appears to be little variability in the fit statistics across conditions, an ANOVA was nonetheless run to compare the effects of parceling strategy and effect size within each of the analytic methods (i.e., Jöreskog-Yang vs. latent variable scores). It should be noted that the structural part of the model is saturated, meaning that the data fit the model perfectly. Discrepancies are due to the measurement models. The basic pattern is the same for all criteria. Effect size has no noticeable impact on any of the fit measures for both the Jöreskog-Yang method and the confirmatory factor analysis used to generate the latent variable scores.

Parceling strategy has a significant effect on model fit within the Jöreskog-Yang method. However, the η^2, a measure of percent of variance attributed to the effect, indicates that parceling strategy accounts for less than 1% of the variance in all of the fit indices except AIC. Parceling strategy accounts for 99% of the variance ($\eta^2 = .99$) in AIC for both the Jöreskog-Yang method and the confirmatory factor analysis. For all effect sizes and analytic methods, the rank order of the AIC by parceling strategy (where smaller is better) is: item > balanced 4 > congeneric 4 > balanced 3 > congeneric 3. This ranking also reflects the rank ordering by number of parameters estimated. Tukey post hocs confirm the significance of the ordering with the exception that balanced 4 is not significantly different from congeneric 4 and balanced 3 is not significantly different from congeneric 3.

With respect to RMSEA, no parceling (i.e., item level analyses) yields significantly lower values than the parceling strategies (i.e., balanced 4, balanced 3, congeneric 4, congeneric 3), $p < .01$, indicating that parceling has a negative impact on the model fit index RMSEA. The effect of parceling on RMSEA within the Jöreskog-Yang method has an $\eta^2 = .01$, and within the latent variable scores method an $\eta^2 = .02$. The results for all model fit outcomes for the latent variable scores follow the same pattern as the Jöreskog-Yang method.

Parameter estimation. The estimated parameter of interest in this study is the structural coefficient for the latent product term. This coefficient is labeled β_3 in Equation 1 (i.e., regression terminology) and γ_3 in Equation A.1 (i.e., structural equation terminology). I will refer to the coefficient as γ_3 to remain consistent with the use of LISREL (Jöreskog & Sörbom, 1996).

The means and standard deviations of γ_3 for each of the study conditions are presented in Table 6a and Table 6b respectively. The simulated 'true' values for each effect size are .125, .15, and .4 respectively. Two basic patterns emerge with respect to the means of γ_3. The balanced parceling strategies tend to overestimate γ_3, and the effect of the overestimation by the balanced strategy is larger for the Jöreskog-Yang method than the latent variable scores method.

Table 5

Fit Indices for Study 2 (averaged across each condition)

Condition	χ^2/df	RMSEA	AIC	NNFI	CFI
.01 JY Item	1.011	0.004	534.532	0.9997	0.9991
.01 JY Bal3	0.964	0.006	110.597	1.0004	0.9993
.01 JY Bal4	0.988	0.006	142.079	1.0001	0.9993
.01 JY Con3	0.957	0.005	110.387	1.0005	0.9993
.01 JY Con4	0.986	0.006	141.964	1.0002	0.9993
.02 JY Item	1.012	0.004	534.767	0.9997	0.9991
.02 JY Bal3	0.977	0.006	111.056	1.0003	0.9992
.02 JY Bal4	0.990	0.006	142.120	1.0001	0.9993
.02 JY Con3	0.974	0.006	110.935	1.0003	0.9993
.02 JY Con4	0.995	0.006	142.395	1.0001	0.9992
.15 JY Item	1.009	0.004	533.630	0.9998	0.9992
.15 JY Bal3	0.980	0.006	111.128	1.0002	0.9993
.15 JY Bal4	0.989	0.006	142.095	1.0001	0.9993
.15 JY Con3	0.964	0.005	110.576	1.0004	0.9994
.15 JY Con4	0.976	0.005	141.377	1.0003	0.9994
.01 LVS Item	1.013	0.004	435.376	0.9997	0.9991
.01 LVS Bal3	0.999	0.007	65.825	1.0000	0.9993
.01 LVS Bal4	1.005	0.007	90.935	1.0000	0.9993
.01 LVS Con3	0.998	0.007	65.816	1.0000	0.9993
.01 LVS Con4	1.014	0.007	91.310	0.9999	0.9993
.02 LVS Item	1.016	0.004	435.983	0.9996	0.9991
.02 LVS Bal3	1.024	0.008	66.447	0.9998	0.9992
.02 LVS Bal4	1.016	0.007	91.393	0.9998	0.9993
.02 LVS Con3	1.012	0.007	66.154	0.9999	0.9993
.02 LVS Con4	1.018	0.007	91.493	0.9998	0.9992
.15 LVS Item	1.011	0.004	434.574	0.9998	0.9992
.15 LVS Bal3	1.015	0.007	66.210	0.9999	0.9993
.15 LVS Bal4	1.008	0.007	91.068	0.9999	0.9993
.15 LVS Con3	1.004	0.007	65.953	1.0000	0.9993
.15 LVS Con4	0.998	0.006	90.684	1.0000	0.9993

Note. Number of replications was 500 in each case. There were no instances of non-convergence or non-admissibility. Effect size is represented by f^2 = .01, .02, and .15. Jöreskog-Yang method with one product indicator is JY. LVS refers to the confirmatory factor analysis used to create the latent variable scores. Balanced parceling strategy is Bal. Congeneric parceling strategy is Con.

Table 6a

Mean of Structural Coefficient for Product Term (γ_3)

	$f^2 = .01$		$f^2 = .02$		$f^2 = .15$	
Parceling Strategy	**JY**	**LVS**	**JY**	**LVS**	**JY**	**LVS**
Item	0.136	0.117	0.163	0.137	0.442	0.371
Balanced3	0.190	0.156	0.224	0.183	0.608	0.495
Balanced4	0.186	0.151	0.218	0.178	0.584	0.481
Congeneric3	0.151	0.126	0.176	0.148	0.481	0.401
Congeneric4	0.147	0.123	0.172	0.144	0.467	0.390

Note. True values are .125, .15, and .4 for f^2 = .01, .02, and .4 respectively. JY = Jöreskog-Yang method. LVS = latent variable score method.

Table 6b

Standard Deviation of Structural Coefficient for Product Term (γ_3)

	$f^2 = .01$		$f^2 = .02$		$f^2 = .15$	
Parceling Strategy	**JY**	**LVS**	**JY**	**LVS**	**JY**	**LVS**
Item	0.073	0.037	0.071	0.037	0.080	0.044
Balanced3	0.082	0.051	0.079	0.052	0.096	0.059
Balanced4	0.086	0.049	0.085	0.050	0.103	0.059
Congeneric3	0.054	0.040	0.053	0.040	0.059	0.044
Congeneric4	0.055	0.039	0.054	0.039	0.061	0.043

Note. The standard deviation of γ_3 is the empirical standard error. JY = Jöreskog-Yang method. LVS = latent variable score method.

The overall shape of the distribution of γ_3 estimates for each of the design cells is depicted in Figures 3a – 3c. The figures show the density curves for the γ_3 estimates across each of the conditions with a vertical dotted line depicting the true population value. The average skewness statistic for γ_3 is .08 (range from -.06 to .33). The only conditions that are significantly skewed are the f^2= .15, Jöreskog-Yang/item, Jöreskog-Yang/congeneric 4, and the latent variable scores/item *method*/parceling combinations. The average kurtosis is .13 (range from -.15 to .58). The only conditions that are significantly leptokurtic are f^2 = .01/latent variable scores/balanced 4 and f^2 =

.02/Jöreskog-Yang/congeneric 4. Table 6b and Figures 3a – 3c show that the Jöreskog-Yang method has greater variability in the estimation of γ_3 than does the latent variable scores method. Within the Jöreskog-Yang method, the congeneric parceling strategies have less variability than either the balanced strategies or the use of items. Less variability equates to more precise estimates of the 'population' parameter.

The standard deviations of γ_3 in Table 6b describe the precision of estimation, but they do not describe the accuracy of estimation. Bias, the average deviation in the estimation of γ_3 from its population parameter determined by the simulation, is presented in Table 7. All but four estimation conditions result in significantly biased results. The congeneric 3 parceling strategy/latent variable scores combination for all effect sizes were not significantly biased, *ps* > .1. The f^2 = .01/congeneric 4/latent variable scores combination also results in no significant bias. Figures 3a – 3c demonstrate bias by the degree of departure of the density curve from the vertical dotted line (i.e., the population parameter determined by the simulation).

Table 7
Bias in Structural Coefficient for Product Term (γ_3)

	f^2 = .01		f^2 = .02		f^2 = .15	
Parceling Strategy	**JY**	**LVS**	**JY**	**LVS**	**JY**	**LVS**
Item	0.011*	-0.008*	0.013*	-0.013*	0.042*	-0.029*
Balanced3	0.065*	0.031*	0.074*	0.033*	0.208*	0.095*
Balanced4	0.061*	0.026*	0.068*	0.028*	0.184*	0.081*
Congeneric3	0.026*	0.001	0.026*	-0.002	0.081*	0.001
Congeneric4	0.022*	-0.002	0.022*	-0.006*	0.067*	-0.010*

Note. Negative values reflect underestimation of γ_3. Positive values reflect overestimation of γ_3. JY = Jöreskog-Yang method. LVS = latent variable score method. * $p < .05$.

The question of how much bias is present across each condition is addressed by ANOVA. A mixed-effects ANOVA where replication is nested within effect size results in statistical significance ($p < .01$) for effect size, method, strategy and each of the two-way and three-way interactions. The three-way interaction accounts for less than 1% of the variance in bias. The η^2 for each of the effects are: effect size = .08, method = .09, parceling strategy = .17, effect size*method = .03, and effect size*strategy = .05, and method*strategy < .01. The pattern of differences was very similar for all effect sizes. Tukey post hoc tests confirm that the balanced 3 and balanced 4 strategies are not significantly different and the congeneric 4 and item strategies are not significantly different, *ps* > .1. All other paired comparisons are different, $p < .05$. That is, both balanced strategies have more bias than either the congeneric strategies or item level analyses. The item strategy has significantly lower mean bias than the congeneric strategies. The congeneric strategies are nearer to zero, whereas the item strategy has negative bias for the latent variable scores method. The congeneric strategies result in less bias than item level for the latent variable scores method, but the reverse is true for the Jöreskog-Yang method.

Another measure of bias is the root mean square error (RMSE), which is a measure of variability akin to the standard deviation in bias. RMSE is the square root of the average squared deviation from the simulated population parameter. Table 8 shows the RMSE for each of the study conditions. For the Jöreskog-Yang method, the balanced strategies have the most variability followed by the item strategy, then the congeneric strategies with the least variability. Using item level indicators (i.e., no parcels) results in an average bias lower than the other parceling strategies, but with an RMSE higher than

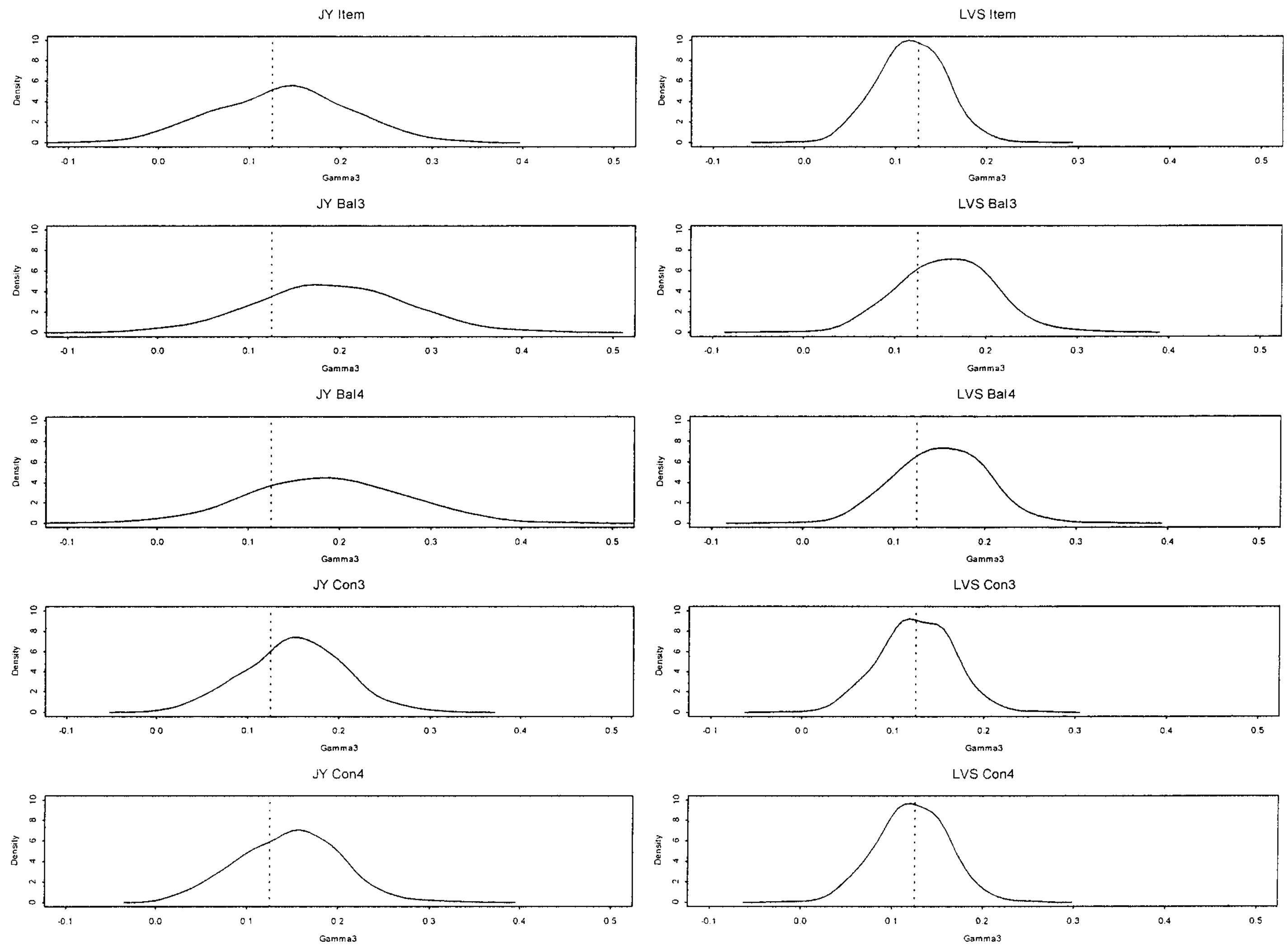

Figure 3a. Distribution of the Structural Coefficient for the Product Term: f^2 =.01.

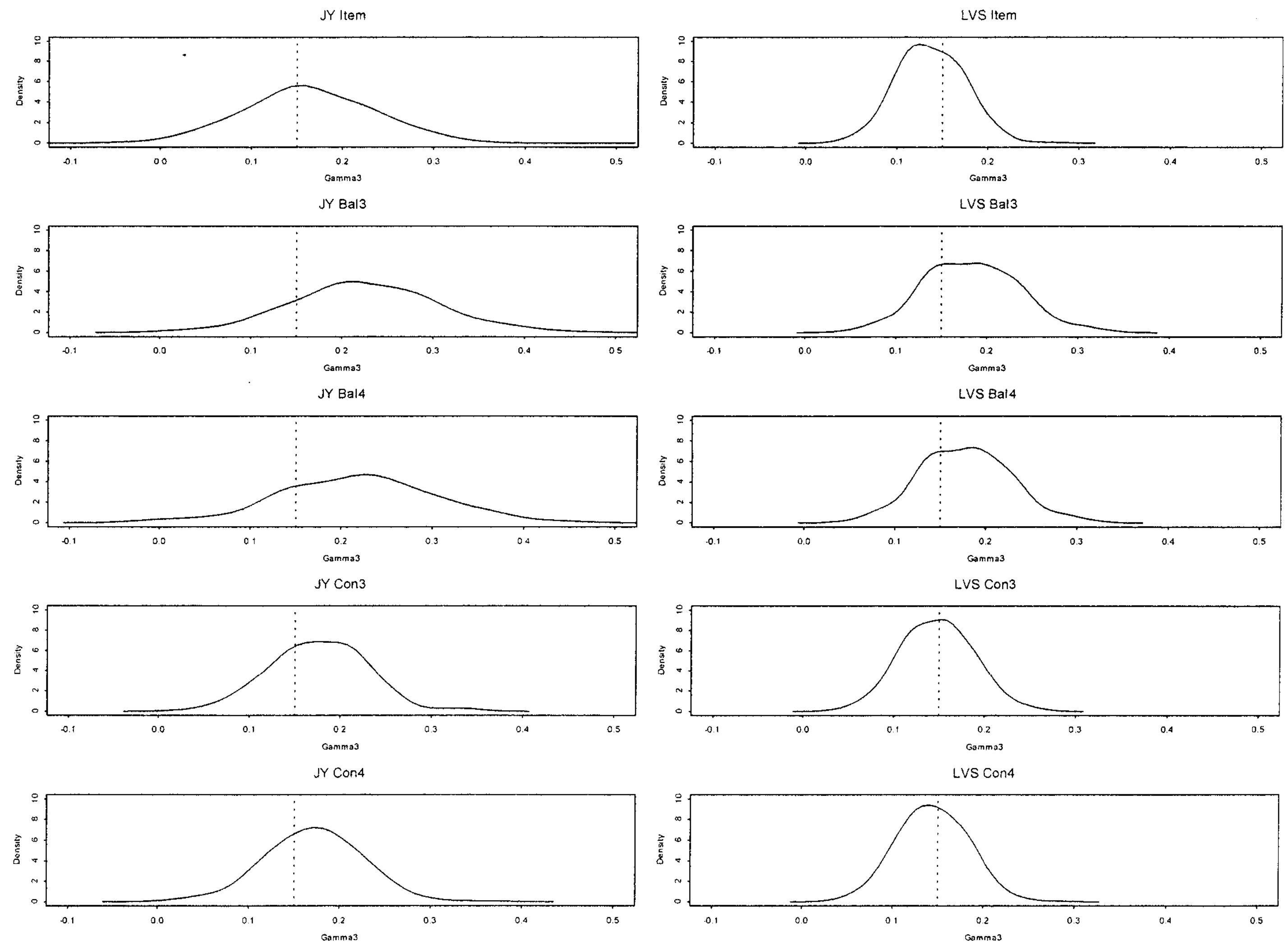

Figure 3b. Distribution of the Structural Coefficient for the Product Term: $f^2 = .02$.

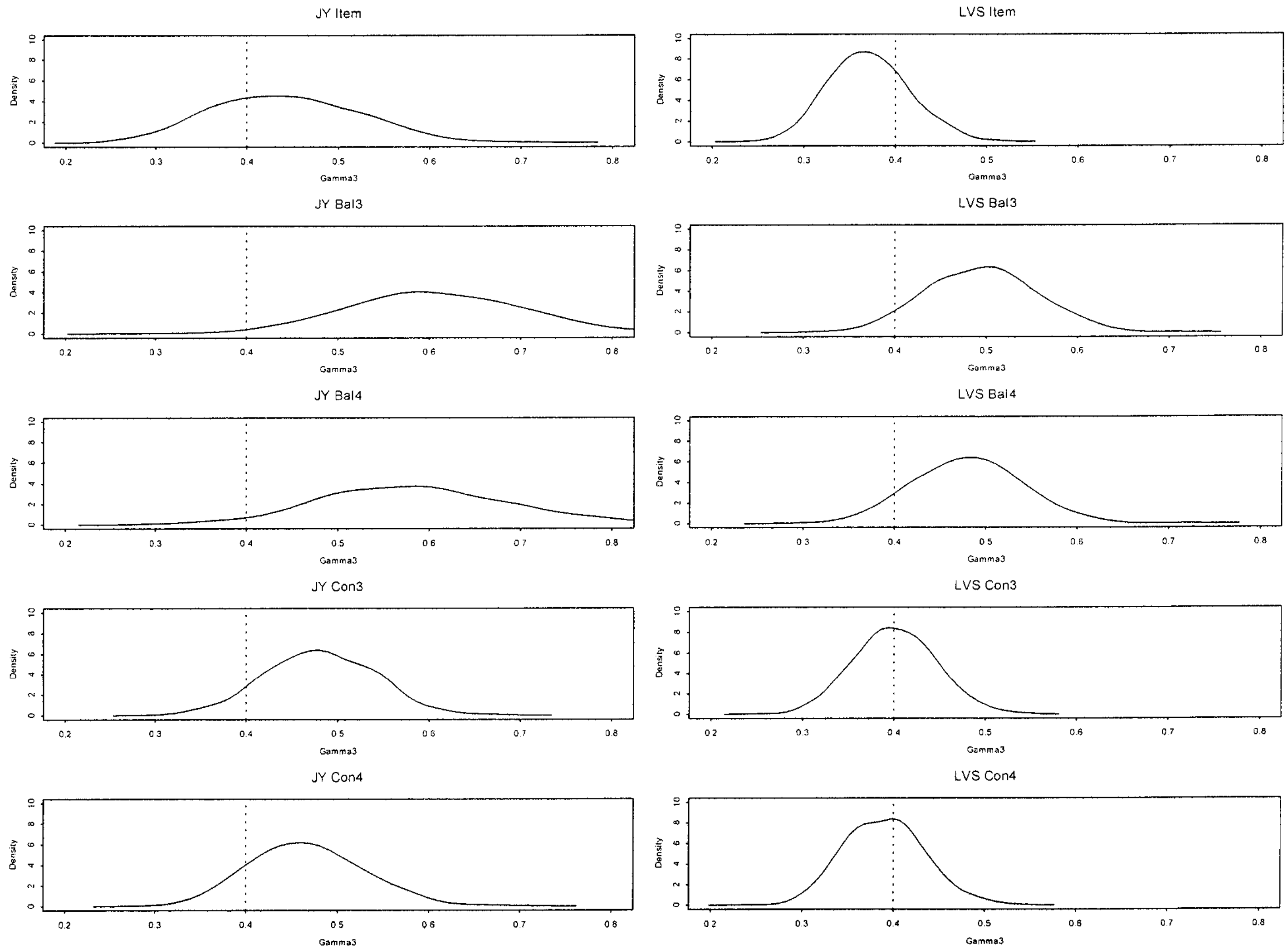

Figure 3c. Distribution of the Structural Coefficient for the Product Term: f^2 =.15.

the congeneric parceling strategies. With respect to the latent variable scores method, the differences in RMSE due to parceling strategy are minimal.

Table 8
Root Mean Square Error in Structural Coefficient for Product Term (γ_3)

	f^2 = .01		f^2 = .02		f^2 = .15	
Parceling Strategy	**JY**	**LVS**	**JY**	**LVS**	**JY**	**LVS**
Item	0.074	0.038	0.073	0.040	0.090	0.052
Balanced3	0.104	0.060	0.109	0.062	0.229	0.112
Balanced4	0.105	0.056	0.109	0.057	0.211	0.100
Congeneric3	0.060	0.040	0.059	0.040	0.100	0.044
Congeneric4	0.059	0.039	0.058	0.039	0.091	0.045

Note. JY = Jöreskog-Yang method. LVS = latent variable score method.

The standard error of γ_3 is directly related to the test of significance for the interaction term. A standard error that is too high leads to lowered power and a standard error that is too low may yield a higher Type I error rate. Tables 9a and 9b show the means and standard deviations of the standard errors for γ_3 for all study conditions. The densities of the standard errors are displayed graphically in Figures 4a – 4c. There are no instances of significant kurtosis. However, nearly all conditions have significantly positive skewness. This is expected since the standard errors should follow a χ^2 distribution. The exceptions are all parceling conditions crossed with the latent variables scores in the smallest effect size, the item parceling strategy with latent variable scores in the largest effect size, and the congeneric 3 parceling strategy crossed with the Jöreskog-Yang method at the smallest effect size. To summarize the results of the mixed effects ANOVA where replication is nested within effect size, the effects have the following η^2: effect size = .01, method = .56, parceling strategy = .25, method*strategy = .07, the remaining effects were all less than .01. In general, the latent variable scores method tends to have lower mean standard errors with less variability than the Jöreskog-Yang method. Within the Jöreskog-Yang method, the parceling strategies have the following pattern of standard errors: congeneric < item < balanced. Tukey post hoc tests comparing parceling strategies within each of the method and effect size conditions yield significant differences for all pairwise comparisons except congeneric 3 vs. congeneric 4.

Table 9a

Mean Standard Error for Structural Coefficient for Product Term (γ_3)

	$f^2 = .01$		$f^2 = .02$		$f^2 = .15$	
Parceling Strategy	**JY**	**LVS**	**JY**	**LVS**	**JY**	**LVS**
Item	0.072	0.038	0.073	0.038	0.080	0.039
Balanced3	0.083	0.052	0.083	0.052	0.092	0.053
Balanced4	0.087	0.050	0.088	0.050	0.097	0.052
Congeneric3	0.055	0.041	0.055	0.041	0.061	0.042
Congeneric4	0.056	0.040	0.056	0.040	0.061	0.041

Note. JY = Jöreskog-Yang method. LVS = latent variable score method.

Table 9b

Standard Deviation of Standard Error for Structural Coefficient for Product Term (γ_3)

	$f^2 = .01$		$f^2 = .02$		$f^2 = .15$	
Parceling Strategy	**JY**	**LVS**	**JY**	**LVS**	**JY**	**LVS**
Item	0.009	0.003	0.009	0.003	0.009	0.003
Balanced3	0.008	0.004	0.008	0.004	0.010	0.004
Balanced4	0.010	0.004	0.010	0.004	0.011	0.004
Congeneric3	0.005	0.003	0.005	0.003	0.005	0.003
Congeneric4	0.005	0.003	0.005	0.003	0.006	0.003

Note. JY = Jöreskog-Yang method. LVS = latent variable score method.

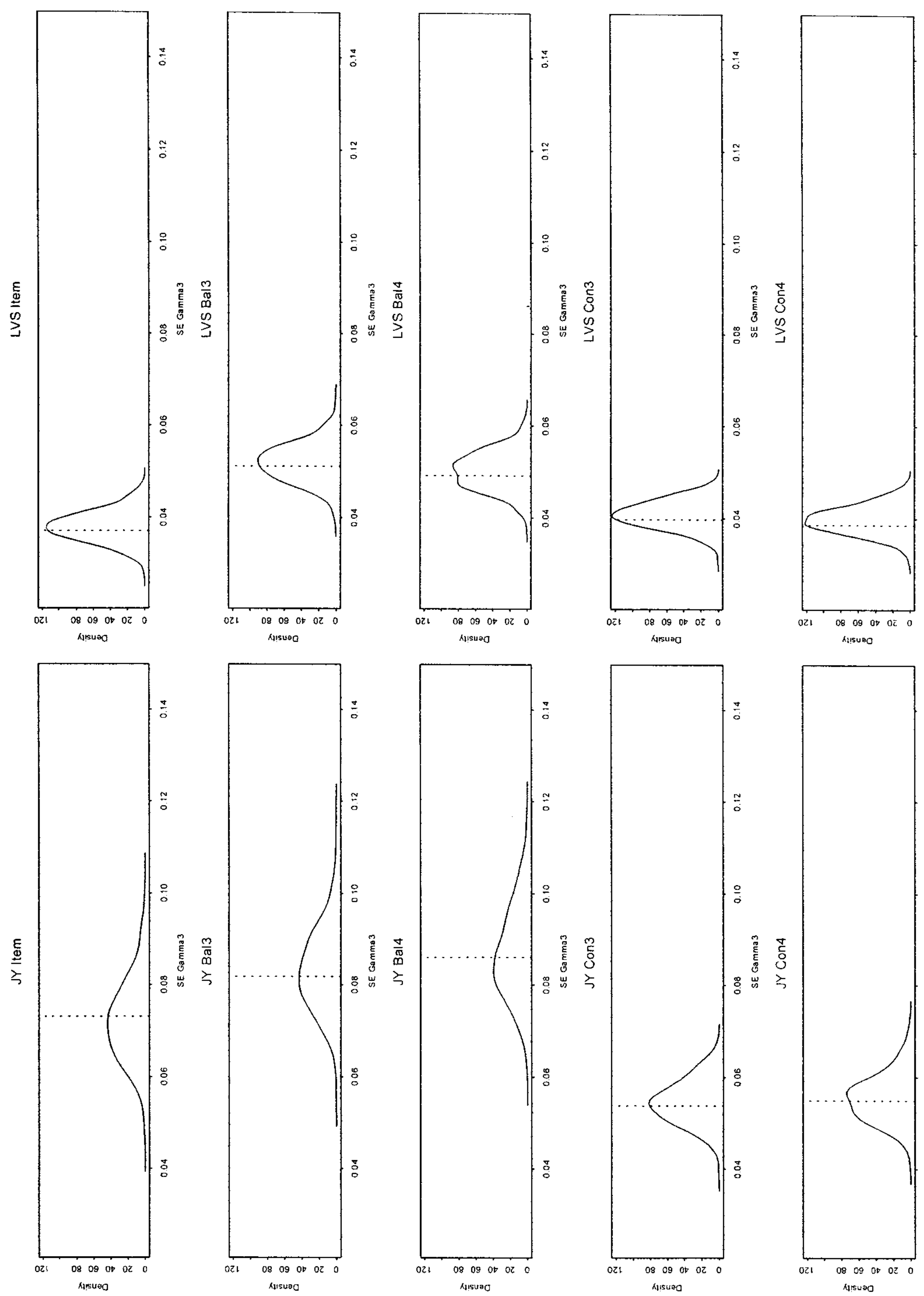

Figure 4a. Distribution of the Standard Errors for Structural Coefficient for the Product Term: f^2=.01.

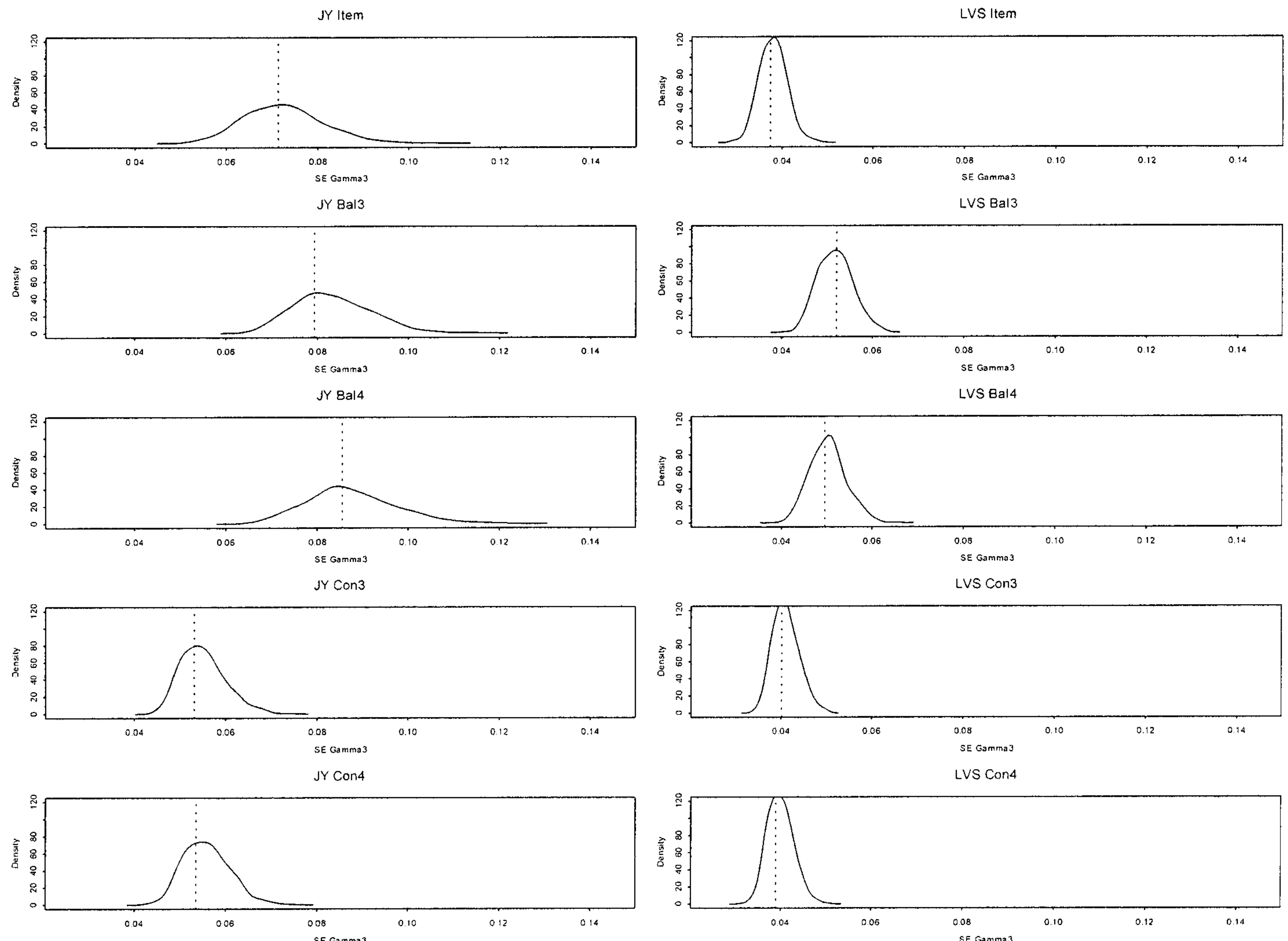

Figure 4b. Distribution of the Standard Errors for Structural Coefficient for the Product Term: f^2=.02.

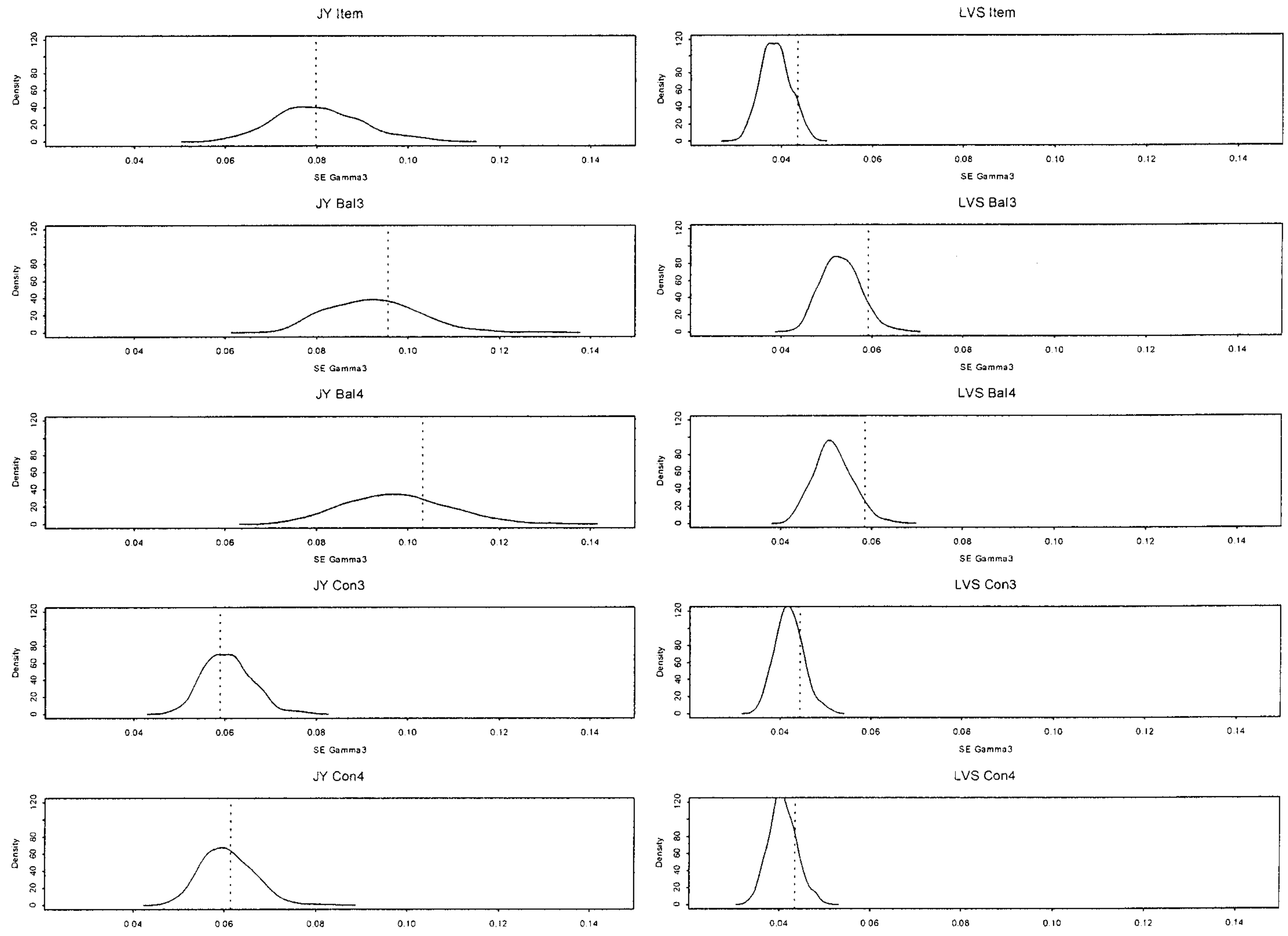

Figure 4c. Distribution of the Standard Errors for Structural Coefficient for the Product Term: f^2=.15.

In general, smaller is better with respect to the standard error, unless the standard error is smaller than the empirical standard error. The empirical standard error is the standard deviation in γ_3. The empirical standard error is depicted in Figures 4a – 4c by the vertical dotted line. The ratio of the estimated standard error to the empirical standard error should be 1.0. Ratios less than or greater than 1.0 reflect under and over estimation, respectively.

The standard error ratios for all conditions are presented in Table 10. In nearly all conditions, the standard errors are biased (i.e., significantly different than 1.0). There are two general patterns that emerge in Table 10 and Figures 4a – 4c. First, there is greater variability in the standard error ratio due to parceling strategy within the Jöreskog-Yang method than the latent variable scores method. Second, variability in the standard error ratio due to parceling strategy is greater at larger effect sizes. These patterns are evident in the boxplots presented in Figure 5. To summarize the results of the mixed effects ANOVA where replication is nested within effect size, the effects have the following η^2: effect size = .11, method = .02, parceling strategy = .01, effect size*method = .04, effect size*strategy = .01, and the remaining effects are all less than .01.

Both the bias in the parameter estimate and its standard error contribute to inferential decision making. If the parameter is estimated to be greater than the population value, and the standard error is estimated to be too small, then the validity of the statistical test (i.e., t – ratio) is in question. That is, the probability of making a Type I error has increased. Likewise, if the parameter estimate is estimated to be too low, and the standard error is too high, then the probability of a Type II error is increased. Table

11a summarizes these possibilities given bias in the parameter estimate and the standard error ratio.

Table 10
Standard Error Ratio of Structural Coefficient for Product Term (γ_3)

	f^2 = .01		f^2 = .02		f^2 = .15	
Parceling Strategy	**JY**	**LVS**	**JY**	**LVS**	**JY**	**LVS**
Item	0.984*	1.026*	1.015*	1.015*	1.007	0.889*
Balanced3	1.015*	1.013*	1.051*	0.994	0.966*	0.898*
Balanced4	1.010	1.021*	1.024*	1.014*	0.945*	0.881*
Congeneric3	1.017*	1.031*	1.036*	1.023*	1.028*	0.947*
Congeneric4	1.013*	1.033*	1.044*	1.027*	0.988*	0.940*

Note. Values above 1.0 indicate overestimation of standard errors. Values below 1.0 indicate underestimation of standard errors. JY = Jöreskog-Yang method. LVS = latent variable score method. *standard error ratio is significantly different than 1.0, $p < .05$.

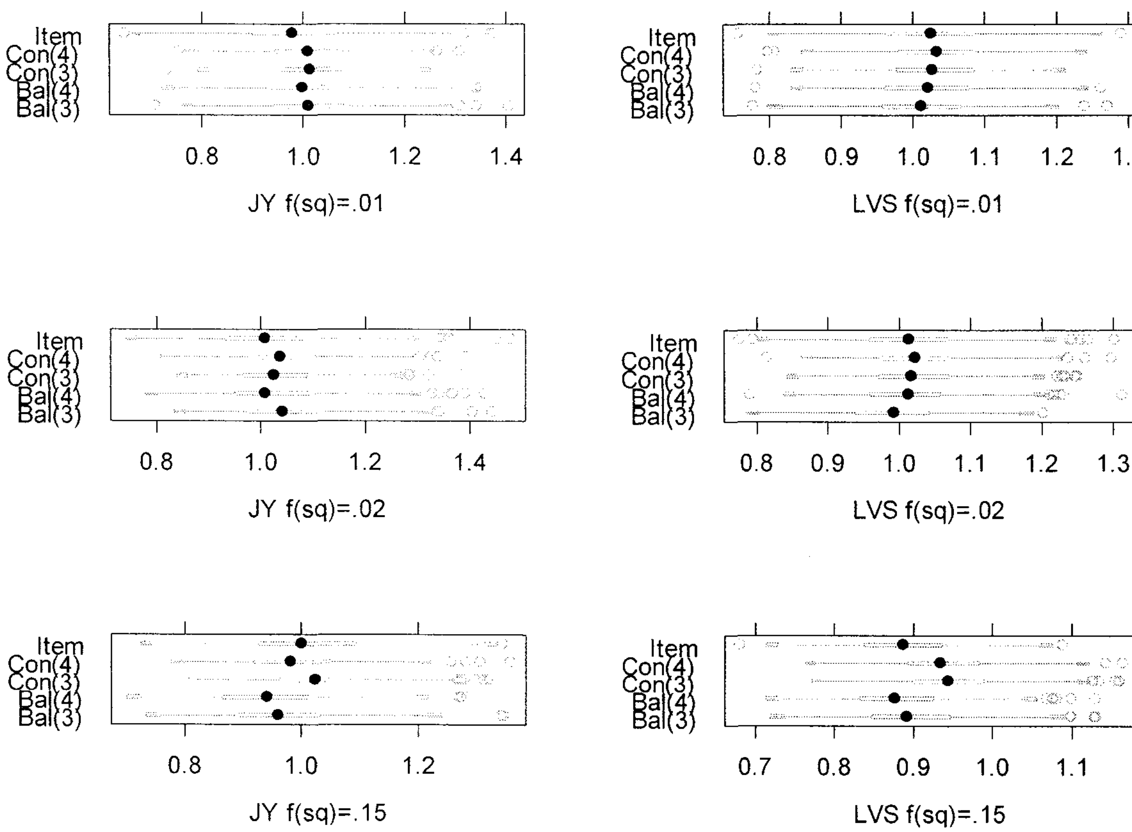

Figure 5. Boxplots of the standard error ratio for all conditions.

In all simulations, the effect was present in the population. There is no possibility to commit a Type I error. However, there is the possibility to commit a Type II error. A Type II error occurs in the current study whenever the *t*-value is less than 1.96. The study conditions are classified according to their bias and standard error ratio in Table 11b. The upper left corner reflects more conservative decisions, whereas the lower right corner reflects less conservative decisions. With only one exception (01JYitem), the conditions classified into the lower right corner are from the larger effect size.

Table 11a
The Potential Effect of Bias and Standard Error Ratio on Statistical Decisions

Standard Error Ratio	Bias in γ_3		
	Negative	0	Positive
> 1	**Type II**	(Type II)	Depends on severity
1	(Type II)	No error	(Type I)
< 1	Depends on severity	(Type I)	**Type I**

Note. Type I and Type II refer to error in decision making.
Bold = increased likelihood, in parentheses = potential for occurrence.

As mentioned above, the *t*-value reflects both the parameter estimate and its standard error. Figures 6a – 6c present the distributions of the *t*-values for all conditions. A vertical, dotted line is plotted at 1.96, the critical value for t, $p < .05$. The portion of the distribution to the left of this dotted line reflects the Type II error rate. That is, this is the proportion of instances where γ_3 was estimated to be non-significant. Because the effect was simulated to exist in the population, non-significance reflects error in decision-making. It is evident in the graphs that the latent variable scores method has a

Table 11b

Classifying Study Conditions based on Bias and Standard Error Ratio for γ_3

Standard Error Ratio	Bias in γ_3		
	Negative	0	Positive
> 1	01LVSitem 02LVSitem 02LVScon4	01LVScon3 01LVScon4 02LVScon3	01JYbal3 01JYcon3 01JYcon4 02JYitem 02JYbal3 02JYbal4 02JYcon3 02JYcon4 15JYcon3 01LVSbal3 01LVSbal4 02LVSbal4
1			01JYbal4 15JYitem 02LVSbal3
< 1	15LVSitem 15LVScon4	15LVScon3	01JYitem 15JYbal3 15JYbal4 15JYcon4 15LVSbal3 15LVSbal4

Note. Effect sizes are labeled 01, 02, and 15. Methods are labeled JY and LVS. Parceling strategy is labeled item, bal3, bal4, con3, and con4. The population effect is present in all simulations; therefore, there are no Type I errors and any non-significant t-value reflects a Type II error. This table is only meant to summarize the likelihood of errors based on the bias and standard error ratio.

substantially lower error rate than does the Jöreskog-Yang method. There is less variability in the estimation of the *t*-values due to parceling strategy in the latent variable scores method than in the Jöreskog-Yang method, and this is more evident at larger effect sizes. Within the Jöreskog-Yang method, the *t*-values appear to be greater for the congeneric parceling strategies than either the balanced or item strategies. Using items as indicators in the Jöreskog-Yang method produces the greatest error rate.

The average effect size estimates, f^2, are presented in Table 12. The latent variable scores method produces little variability in f^2 due to parceling strategy. However, there is great variability in f^2 due to parceling strategy within the Jöreskog-Yang method. To assist in interpretation of the method by strategy interaction, an interaction plot is graphed in Figure 7. A horizontal line represents the expected effect size, $f^2 = .01$, .02, and .15. For the lowest effect size, the latent variable scores method overestimates the effect size and the congeneric methods within the Jöreskog-Yang method accurately estimate the f^2. For the larger effect sizes, $f^2 = .02$ and .15, both the latent variable scores and the Jöreskog-Yang methods underestimate the effect size. For all effect sizes, the congeneric strategies outperform item or balanced strategies when utilizing the Jöreskog-Yang method in determining the population effect size.

Table 12

Mean Effect Size (f^2) for Product Term (γ_3) for Study 2

	$f^2 = .01$		$f^2 = .02$		$f^2 = .15$	
Parceling Strategy	**JY**	**LVS**	**JY**	**LVS**	**JY**	**LVS**
Item	0.006	0.013	0.008	0.018	0.039	0.118
Balanced3	0.008	0.013	0.010	0.017	0.056	0.112
Balanced4	0.007	0.013	0.009	0.017	0.047	0.112
Congeneric3	0.011	0.013	0.014	0.018	0.081	0.117
Congeneric4	0.010	0.013	0.013	0.018	0.076	0.118

Note. JY = Jöreskog-Yang method. LVS = latent variable score method.

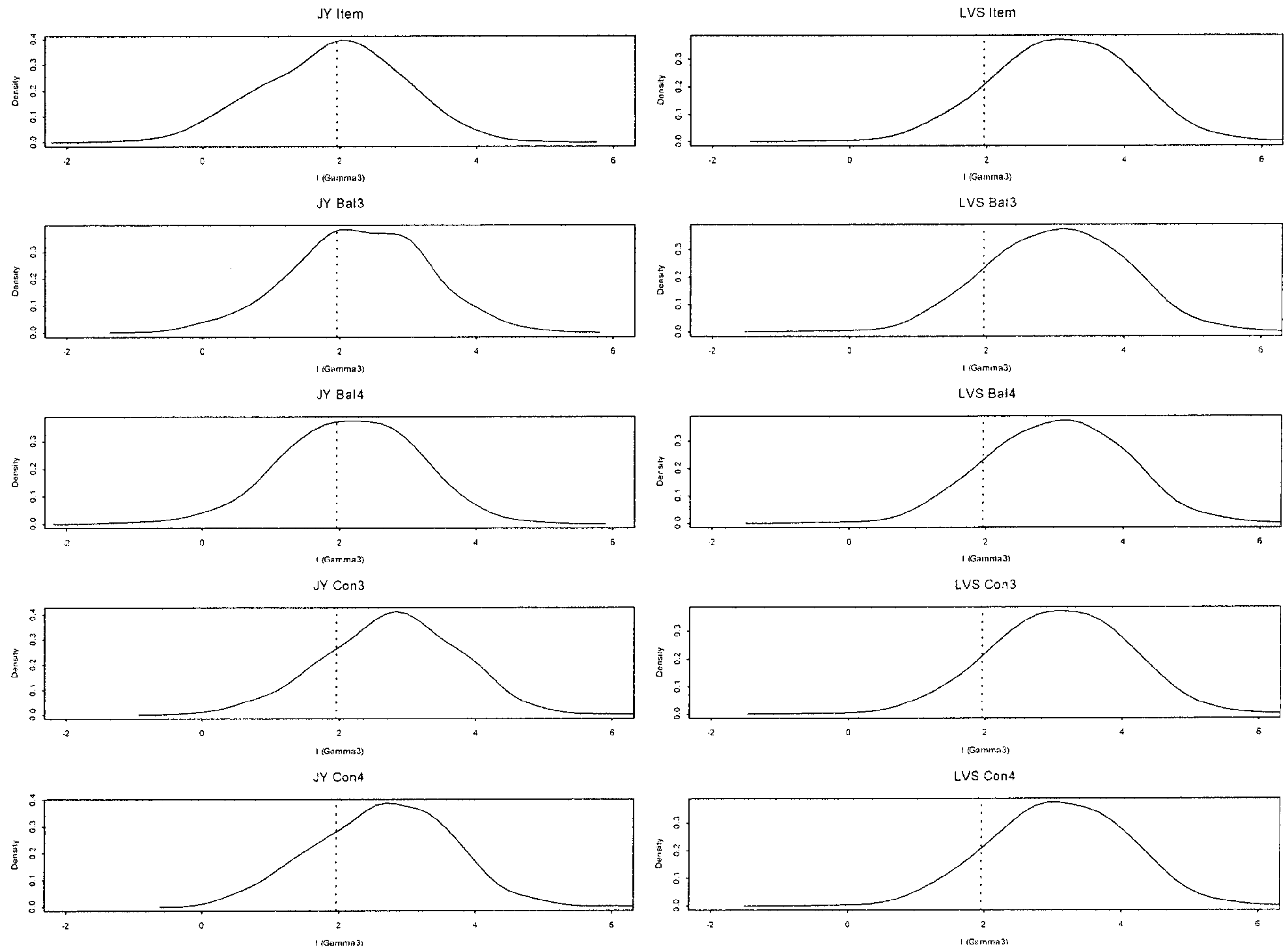

Figure 6a. Distribution of the t-values for the Structural Coefficient for the Product Term: f^2=.01. Vertical dotted line is 1.96, the cut-off for $p < .05$.

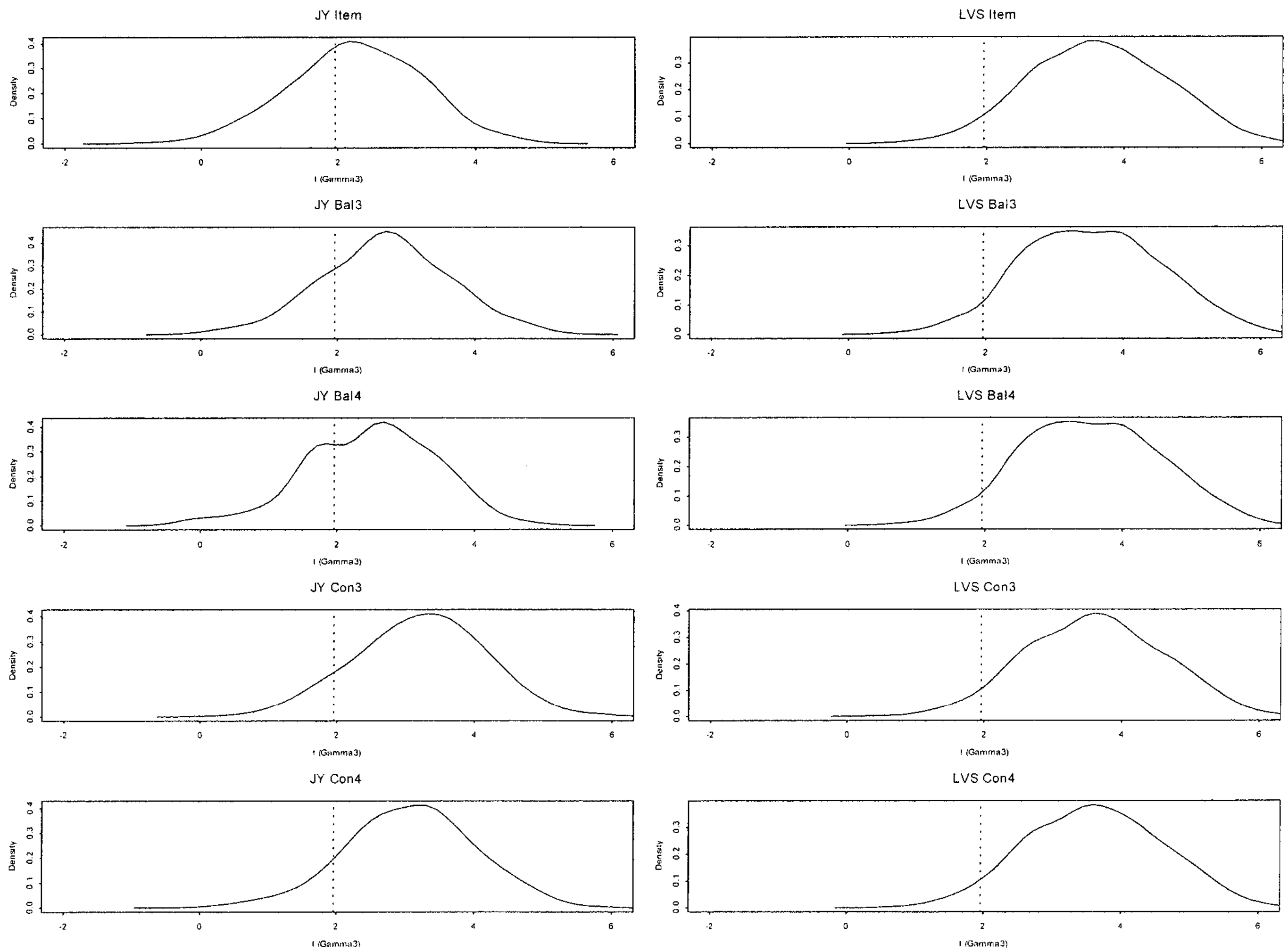

Figure 6b. Distribution of the t-values for the Structural Coefficient for the Product Term: f^2=.02. Vertical dotted line is 1.96, the cut-off for $p < .05$.

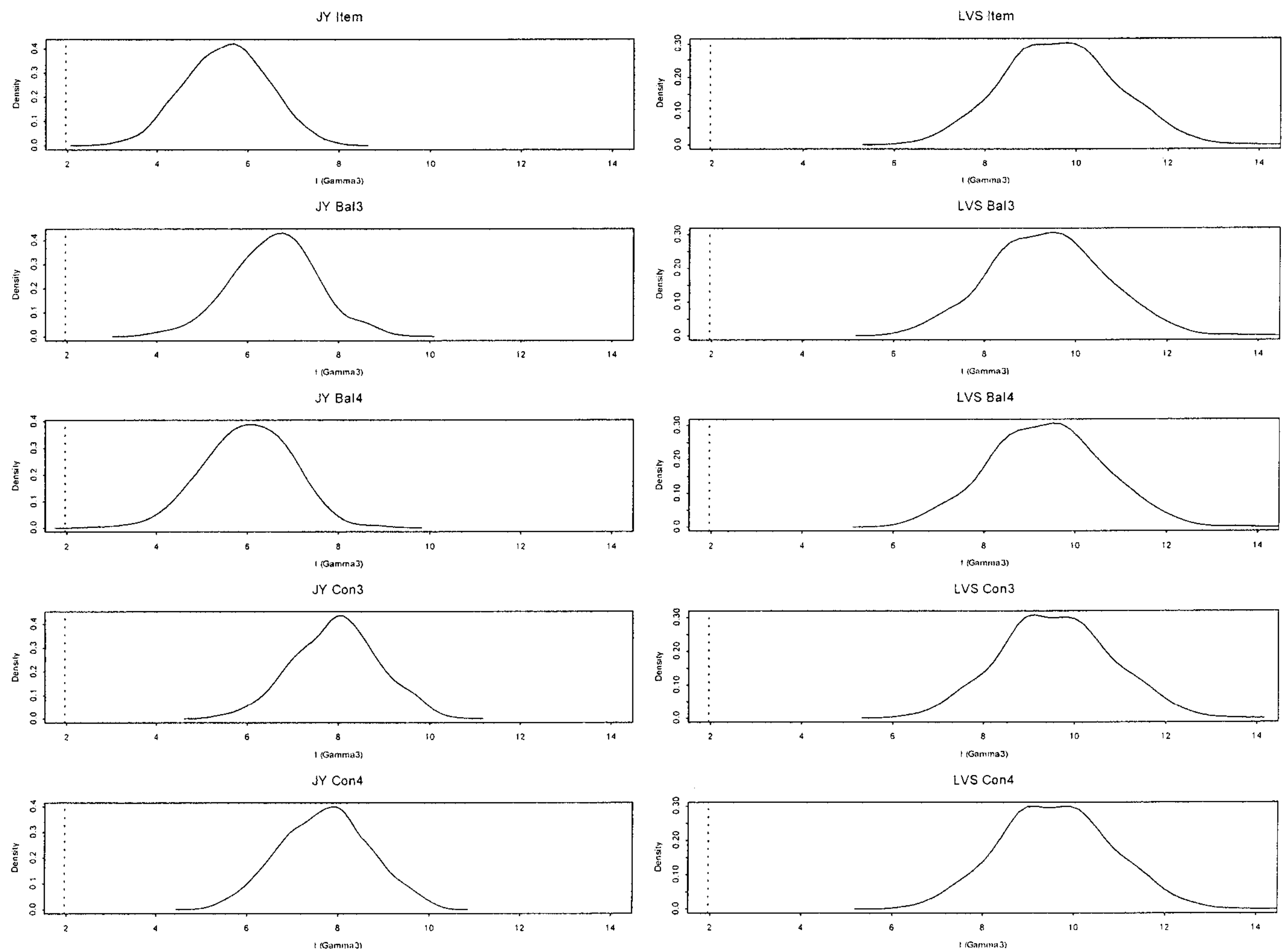

Figure 6c. Distribution of the t-values for the Structural Coefficient for the Product Term: f^2=.15. Vertical dotted line is 1.96, the cut-off for $p < .05$.

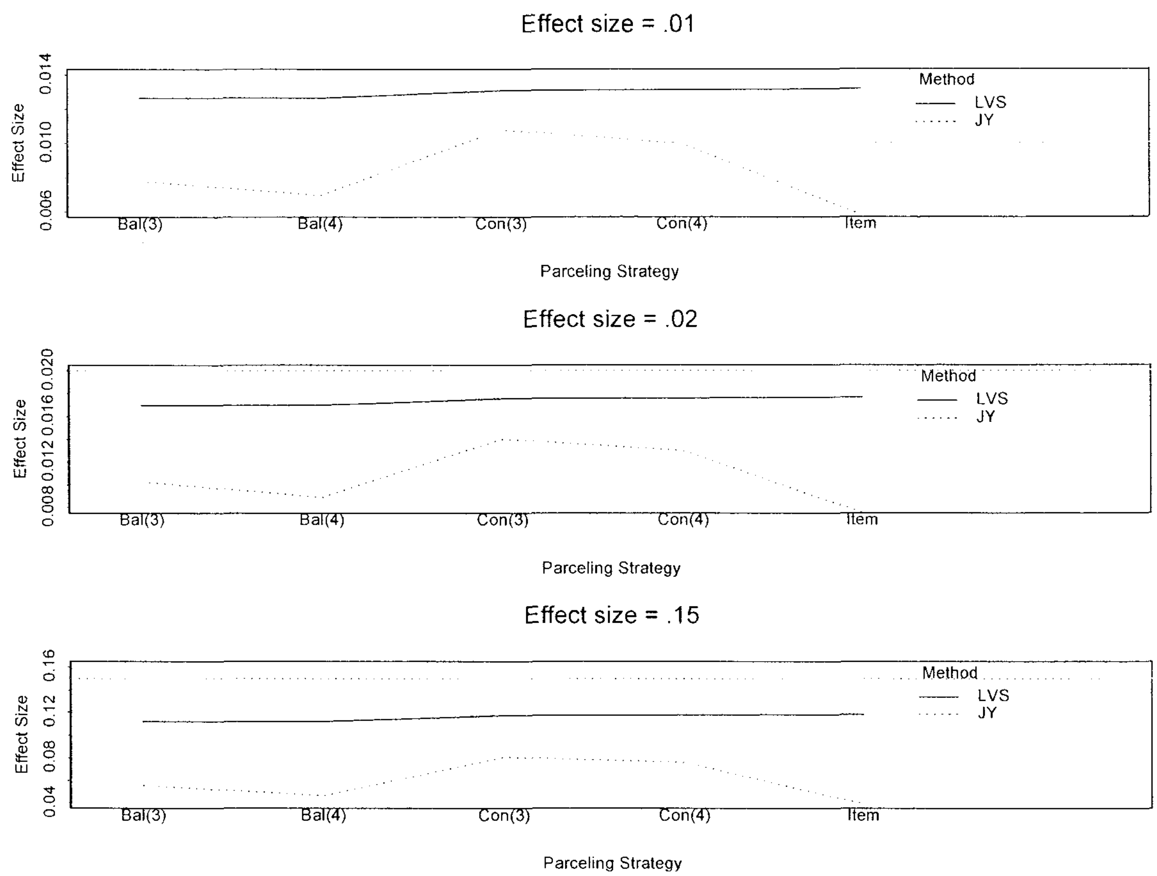

Figure 7. Interaction Plot of the Effect Size index f^2.

Summary. The following general statements summarize the results of Study 2:

- All estimation conditions lead to acceptable model fits.
- All but four estimation conditions result in significantly biased results. These four conditions are the congeneric3 with latent variable scores for all three effect sizes and the congeneric4 with latent variable scores for the f^2 = .01 effect size.
- There is greater variability in the estimation of γ_3 with the Jöreskog-Yang method than the latent variable scores method.
- For either analytic method, the balanced parceling strategies lead to the least accurate and least precise estimates of γ_3.
- Within the Jöreskog-Yang method, the congeneric parceling strategies lead to the most accurate and most precise estimates of γ_3.
- Within the latent variable scores method, the differences in RMSE due to parceling strategy are minimal.

- The latent variable scores method is more precise in the estimation of the standard errors of γ_3 than the Jöreskog-Yang method.
- There is greater variability in the standard error ratio due to parceling strategy within the Jöreskog-Yang method than the latent variable scores method.
- Variability in the standard error ratio due to parceling strategy is greater at larger effect sizes.
- There is greater variability in the estimation of the *t*-values due to parceling strategy in the Jöreskog-Yang method than in the latent variable scores method.
- The latent variable scores method is superior to the Jöreskog-Yang method with respect to statistical decision-making (i.e., fewer Type II errors).
- Within the Jöreskog-Yang method, the *t*-values appear to be greater for the congeneric parceling strategies than either the balanced or item strategies.
- Using items as indicators in the Jöreskog-Yang method produces the greatest error rate.

GENERAL DISCUSSION AND CONCLUSION

It is well known that measurement error greatly complicates the detection of statistical interactions in applied psychology. There are a number of methods for controlling the effects of measurement error when investigating interactive relationships (see Cortina et al., 2001 and Schumacker & Marcoulides, 1998 for reviews). The leading, or most elegant of these approaches is the Kenny-Judd method as advanced by Jöreskog and Yang (1996). The problems with the Jöreskog-Yang method, however, are model complexity and bias in parameter estimates. The purpose of this research was to investigate two potential solutions to these problems in controlling for measurement error when assessing interactions: the use of parcels and the use of latent variable scores.

The purpose of Study 1 was to (1) use research data to test the multiplicative relationship of competitive climate and trait competitiveness on affective and instrumental support from coworkers, and (2) assess the potential variability in estimation procedures. A significant interaction effect indicates that the relationship of climate on the support variables is contingent upon individual differences in competitiveness.

For affective support, five of six conditions resulted in statistical significance. The effect size of the interaction was variable for the conditions. The Jöreskog-Yang method yielded lower effect sizes than the latent variable scores method. Parceling had the following effects on effect size estimation: the balanced strategy was lower than the item strategy, which was lower than the congeneric strategy. The combination of latent variable scores with the congeneric parceling strategy yielded the highest effect size estimate. Based on Figure 1, the form of the interaction was quite similar for all conditions despite variability in effect size estimates.

For instrumental support only three of six conditions resulted in statistical significance for the interaction effect. The effect sizes were too low for the sample size to have adequate power to reach statistical significance. The choice of method and strategy would have resulted in different conclusions as to the relationship of perceptions of a competitive climate interacting with trait competitiveness in influencing perceptions of instrumental support from coworkers. The largest effect size was estimated with the Jöreskog-Yang procedure with the congeneric parceling strategy. The weakest effect size was estimated with the Jöreskog-Yang method with the balanced parceling strategy. Parceling makes a difference, but the difference depends on which strategy is used. Figure 2 demonstrates that the form of the interaction is also variable depending on estimation condition. The Jöreskog-Yang method coupled with the congeneric parceling strategy indicates a disordinal interaction, whereas the same method with a balanced parceling strategy indicates no interaction. The latent variable scores method coupled with the congeneric parceling strategy indicates an ordinal interaction.

The purpose of Study 2 was to simulate data with known population values and compare each of the analytic methods and parceling strategies for each of three independent effect sizes in retrieving those values. The general question was whether the use of latent variable scores differs substantially from the Jöreskog-Yang method, and whether the use of parcels can improve the estimation of the interaction effects. The answer to both of these questions is a complex yes.

In this study, I confirmed what others have already noted. The Jöreskog-Yang method produces biased parameter estimates of the interaction effect. The Jöreskog-Yang method has great variability in estimating population parameters. Both the empirical standard error and the estimated standard errors are quite variable. The use of parcels, in particular parcels created via the congeneric parceling strategy, improved the estimation of the interaction effect. In contrast, the use of the balanced parceling strategy worsened the estimation of the interaction effect. The congeneric strategy resulted in more accurate and precise estimation of the population parameters in comparison to the use of either items or the balanced parceling strategy. Parceling can improve model estimation if done properly.

Parceling had less of an effect when the latent variable scores method was utilized. However, the latent variable scores method, for all parceling strategies, yielded superior estimation of the parameter estimates, estimation of the effect sizes, and statistical decisions in contrast with the Jöreskog-Yang method. Only when the congeneric strategy was used in the Jöreskog-Yang method were the results comparable to the latent variable scores method.

With respect to effect size, the methods were less accurate and less precise at the larger effect size (e.g., $f^2 = .15$). In hindsight this makes sense. Measurement error has the greatest impact on model estimation for the product term. When the product term or interaction effect size is large, the product, not the component terms, is driving the relationship. At larger effect sizes, measurement error has the greatest capacity to interfere with proper estimation. Hence, there is greater variability, less accuracy and precision, at larger effect sizes. This indicates that choice of method or strategy may have less of an impact on the estimation of the interaction effect size when the population effect size is rather small (e.g., less than .02). That being said, Figure 6a shows that at $f^2 = .01$, the Jöreskog-Yang method with the use of the product of a single item from each component term as the indicator for the latent product term, is as likely to accept the null hypothesis as it is to reject it.

With respect to Study 1, the results of Study 2 suggest that the results from the latent variable scores are the most plausible. With respect to the use of parcels, either the use of items or the congeneric parceling strategy produced plausible results whereas the results from the use of the balanced strategy are questionable. Therefore, the presence of the interaction is tenable since the use of the congeneric strategy with either the latent variable score method or the Jöreskog-Yang method lead to significant results. The f^2s for affective and instrumental support are likely .019 and .005 respectively. This conclusion is from the use of the congeneric parceling strategy in conjunction with the latent variable score method.

Practical Implications

The results of this dissertation may have a number of practical implications. First, latent variable scores are relatively easy to compute. The Kenny-Judd model went largely unused for over a decade because of the sophisticated nature of the nonlinear constraints necessary for model estimation. Researchers may be more inclined to test theories with interaction effects given the simpler latent variable scores method. The use of latent variable scores also has the byproduct of forcing a separation of measurement model estimation from structural model estimation. Separately estimating the measurement model from the structural model will assist researchers in identifying the sources of problems with their models. The use of latent variable scores may also provide an alternative to the estimation of multilevel models by creating the latent variable scores and inputting these data in alternate statistical packages (e.g., Hierarchical Linear Modeling).

Second, if a single full-information model is to be estimated, the congeneric parceling strategy demonstrated promise in improving the estimation of the interaction effect when using the Jöreskog-Yang method. Although not directly tested, it seems unlikely that the Jöreskog-Yang method in conjunction with the congeneric parceling strategy would lead to increases in Type I error. The Jöreskog-Yang method may also be useful in estimating other types of models such as those with quadratic terms, when the congeneric parceling strategy is used. Further, parceling can reduce model complexity when more than three latent predictors are used. Each of these last points is fruit for future research.

Limitations and Future Research

As with all Monte Carlo research, generalizations are limited to the design of the study. As a first step at addressing the research questions in this dissertation, the data were simulated based on normality. Reliability of the items was considered but not systematically manipulated. Further, the correlation between the latent predictors was held constant at $r = .3$. I acknowledge that future research could manipulate each of these variables.

Several future investigations come to mind. As mention previously, the data generated were normal and continuous. However, most research in the behavioral and organizational sciences relies on data that are collected using Likert-type scales. Individual responses to such scales are typically non-normal and ordinal. The extent to which the present study's findings generalize to such data is not yet known. Future research should include investigations of (1) ordinal data rather than continuous, and (2) various degrees of non-normality in the component terms. For instance, the use of items as a parceling strategy in the current design fared well, but may not do so compared to parcels when the items are not continuous.

While three effect sizes were investigated, the form of the interaction ultimately remained the same in this study. It is not known how the form of the interaction may be related to the detection of the interaction via the procedures delineated here. For instance, does the presence of a disordinal interaction exacerbate the bias in parameter estimates when utilizing the Jöreskog-Yang method? Such a question is an area for future research inquiry.

As with the previous research on the Kenny-Judd model, the present study investigated a very simple relationship: $Y = X + Z + XZ$. Future research should address the effects of more complex models. For instance, does the presence of covariates in the model attenuate the ability to detect interactions? Jaccard and Wan (1995) introduced covariates into their model but did not explain their effect on the detection of the interaction. In more complex models, moderators can also act as mediators. Other multiplicative relationships exist such as quadratic functions (i.e., squared terms). Future research could investigate these more complex relationships. To try and address these questions in the current research would have made it unmanageable.

I did not directly address the potential for making a Type I error. This is of concern because of the positive bias in many of the study conditions. The question is whether the nominal Type I error rate of 5% is reflected in the empirical distributions of statistical tests when no interaction effect is present. From the distribution of t-values, it can be inferred that Type I errors are not a problem, but this should be directly assessed in future research.

One final concern that I will address is that of the potential for the rank ordering of items to vary across replications. Given the measurement parameters (see Table 4), it is possible that some items in some replications will not have the exact ranking as specified by the population algorithm. This has direct implications on the item-to-construct-balance parceling strategy. A sample of 20 of the 500 replications were examined to see the extent to which miss ranking occurs. Of the 40 sets of factor loadings (i.e., 20 for *X* and 20 for *Z*), only seven had absolute rankings in the correct order. The balanced strategy with three parcels was further scrutinized. For parcel 1, the reference indicator used to form the product indicator, only 17 out of 40 sets of items were assigned correctly. The majority of mis-assigned items involved only one item. To understand the extent to which mis-assigned items might have influenced the Study 2 results, the parcels were specified for each of the 20 replications according to the empirical loadings. That is, for each replication, the items were parceled separately and the new interaction indicator was formed. The Jöreskog-Yang method was then applied to these 20 replications and compared to the previously obtained results. The correlation between the obtained *t*-values for γ_3 from the empirically assigned balanced parcels and the algorithm assigned balanced parcels is $r = .92$, $p < .05$. The *t*-values include both the estimated γ_3 coefficient and the standard error. Although some items were mis-parceled according the balanced strategy, this did not lead to a serious threat to the results for the Jöreskog-Yang method.

Conclusion

The choice of analytic method and parceling strategy do have implications for the detection of statistical interactions in the presence of measurement error. Although the Jöreskog-Yang procedure does tend to overestimate the product coefficient, the degree of overestimation can be attenuated through the use of parcels. However, a congeneric approach to parceling is preferred over a balanced approach. Measurement error must be isolated rather than distributed in order to improve estimation.

The use of latent variable scores can greatly improve upon the precision of the estimation of interactions. However, many questions remain and no single method was 100% accurate in the estimation. With respect to Study 1, decisions regarding reward structures (i.e., factors influencing a competitive climate) would be misinformed if not accounting for individual levels of trait competitiveness. The interactive effects were more pronounced on affective support rather than instrumental support. Nonetheless, such theories with contingencies require analytic methods that can accurately estimate the effect sizes of the interaction to be useful in applied settings.

REFERENCES

Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York: Guilford.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park: Sage Publications.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice Hall.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317-332.

Bagozzi, R. P., & Edwards, J. E. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1,* 45-87.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9,* 78-102.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum.

Barrett, P. T., & Kline, P. (1981). Radial parcel factor analysis. *Personality and Individual Differences, 2,* 311-318.

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 88,* 238-246.

Bollen, K. (1989). *Structural equations with latent variables.* NewYork: John Wiley.

Bollen, K. (1995). Structural equation models that are nonlinear in latent variables: A least squares estimator. In P. M. Marsden (Ed.), *Sociological methodology*, (pp. 223-251). Cambridge, MA: Blackwell.

Bollen, K. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika, 61,* 109-121.

Bollen, K., & Paxton, P. (1998). Two-stage least squares estimation of interaction effects. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation models* (pp. 125–151). Mahwah, NJ: Lawrence Erlbaum .

Borhnstedt, G. W., & Goldberger, A. S. (1969). On the exact covariance of products of random variables. *Journal of the American Statistical Association, 64,* 1439-1442.

Borhnstedt, G. W., & Marwell, G. (1978). The reliability of products of two random variables. *Sociological Methodology, 9,* 254-273.

Brown, S. P., Cron, W. L., & Slocum, J. W., Jr. (1998). Effects of trait competitiveness and perceived intraorganizational competition on salesperson goal setting and performance. *Journal of Marketing, 62,* 88-98.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin, 93,* 549-562.

Cattell, R. B. (1974). Radial item parcel factoring vs. item factoring in defining personality structure in questionnaires: Theory and experimental checks. *Australian Journal of Psychology, 26,* 103-119.

Cattell, R. B., & Bursdal, C. A. (1975). The radial parcel double factor design: A solution to the item-versus-parcel controversy. *Multivariate Behavioral Research, 10,* 165-179.

Champoux, J. E., & Peters, W. S. (1987). Form, effect size and power in moderated regression analysis. *Journal of Occupational Psychology, 60,* 243-255.

Chaplin, W. F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality, 59,* 143-178.

Chou, C. P., Bentler, P., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44,* 347-357.

Cleveland, W. S. (1993). *Visualizing data.* Summit, NJ: Hobart Press.

Cohen, J. (1962). The statistical power of abnormal – social psychological research: A review. *Journal of Abnormal and Social Psychology, 65,* 145-153.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2^{nd} ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Comrey, A. L. (1970). *Manual for the Comrey Personality Scales.* San Diego: Educational and Industrial Testing Service.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98-104.

Cortina, J. M. (2002). Big things have small beginnings: An assortment of "minor" methodological misunderstandings. *Journal of Management, 28,* 339-362.

Cortina, J. M., Chen, G., & Dunlap, W. P. (2001). Testing interaction effects in LISREL: Examination and illustration of available procedures. *Organizational Research Methods, 4,* 324-360.

Cortina, J. M., & DeShon, R. P. (1998). Determining relative importance of predictors with the observational design. *Journal of Applied Psychology, 83,* 798-804.

Ducharme, L. J., & Martin, J. K. (2000). Unrewarding work, coworker support, and job satisfaction: A test of the buffering hypothesis. *Work and Occupations,* 27, 223-243.

Endler, N. S. & Magnusson, D. (Eds.). (1976). *Interactional psychology and personality.* New York: Hemisphere.

Hagtvet, K. A., & Nasser, F. M. (2004). How well do item parcels represent conceptually defined latent constructs? A two-facet approach. *Structural Equation Modeling, 11,* 168-193.

Hall, R. J., Snell, A. F., & Singer Foust, M. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods, 2,* 233-256.

Helmreich, R. L., & Spence, J.T. (1978). The work and family orientation questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. *JSAS Catalog of Selected Documents in Psychology, 8,* 35.

House, R. J. (1971). A path-goal theory of leader effectiveness. *Administrative Science Quarterly, 16,* 321-339.

House, R. J. (1996). Path-goal theory of leadership: Lessons, legacy, and a reformulated theory. *Leadership Quarterly, 7,* 323-352.

Hu, L. T., Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Hu, L. T., Bentler, P., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351-362.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression (2^{nd} ed.).* Thousand Oaks: Sage Publications.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newbury Park: Sage Publications.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin, 117,* 348-357.

Jaccard, J., & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression.* Thousand Oaks: Sage Publications.

Johnson, D. W., & Johnson, R. (1989). *Cooperation and competition: Theory and research.* Edina, MN: Interaction Books.

Jöreskog, K. G. (1998). Interaction and nonlinear modeling: Issues and approaches. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation models* (pp. 239-250). Mahwah, NJ: Lawrence Erlbaum.

Jöreskog, K. G. (2000, July 10). *Latent variable scores and their uses.* Retrieved July 11, 2004, from http://www.ssicentral.com/lisrel/column6.pdf

Jöreskog & Sörbom (1996a). *LISREL 8: User's reference guide.* Chicago: Scientific Software International.

Jöreskog & Sörbom (1996b). *PRELIS 2: User's reference guide.* Chicago: Scientific Software International.

Jöreskog, K. G., Sörbom, D., du Toit, S., & du Toit, M. (2000). *LISREL 8: New statistical features*. Chicago: Scientific Software International.

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 57-88). Hillsdale, NJ: Lawrence Erlbaum.

Kendall, M. G., & Stuart, A. (1958). *The advanced theory of statistics* (Vol. 1). London: Griffin.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96,* 201-210.

Kerr, S. & Jermier, J. M. (1978). Substitutes for leadership: Their meaning and measurement. *Organizational Behavior and Human Performance, 22,* 375-403.

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement, 54,* 757-765.

Kozlowski, S. W. J., & Bell, B.S. (2003). Work groups and teams in organizations. In W.C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (vol. 12, pp. 333-375). Hoboken, NJ: John Wiley & Sons.

Lawley, D. N. & Maxwell, A. E. (1971). *Factor analysis as a statistical method*, (2nd edition). London: Butterworths.

Lee, S., Song, X., & Poon, W. (2004). Comparison of approaches in estimating interaction and quadratic effects of latent variables. *Multivariate Behavioral Research, 39,* 34-67.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9,* 151-173.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4,* 192-211.

Major, D. A., Davis, D. D., Sanchez-Hucles, J. V., Mann, J. E., Clarke, S. M., Cardenas, R. A., Fletcher, T. D., & Germano, L. M., (2003). *Annual Report: Climate for Opportunity and Inclusion: Improving the Recruitment, Retention and Advancement of Women and Minorities in IT.* Technical Report prepared for National Science Foundation (Grant No. EIA-0204430). Norfolk, VA: Old Dominion University.

Marsh, H. W., Hau, K., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181-220.

Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1992). Influences of individual and situational influences on training effectiveness. *Academy of Management Journal, 35,* 828-847.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376-390.

McDonald, R. P., & Ringo Ho, M.-H. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7,* 64-82.

Mooney, C. Z. (1997). *Monte Carlo simulation.* Thousand Oaks: Sage Publications.

Nasser, F., & Wisenbaker, J. (2003). A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement, 63,* 729-757.

Pervin, L. A. (1987). Person-environment congruence in the light of the person-situation controversy. *Journal of Vocational Behavior, 31,* 222-230.

Pervin, L. A. (1989). Persons, situations, interactions: The history of a controversy and a discussion of theoretical models. *Academy of Management Review, 14,* 350-360.

Pervin, L. A., & Lewis, M. (Eds.). (1978). *Perspectives in interactional psychology.* New York: Plenum.

Ping, R. A., Jr. (1995). A parsimonious technique for interaction and quadratic latent variables. *Journal of Marketing Research, 32,* 336-347.

Ping, R. A., Jr. (1996). Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin, 119,* 166-175.

Ross, L. & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* Philadelphia: Temple University Press.

Sattora, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structural analysis. *Proceeding of the Business and Economic Statistics Section of the American Statistical Association*, 308-313.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Journal of Applied Psychology, 8,* 350-353.

Schneider, B. (1983). Interactional psychology and organizational behavior. In L. Cummings & B. Staw (Eds.), *Research in organizational behavior,* (Vol. 5, pp. 1-31), Greenwich, CT: JAI Press.

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40,* 437-453.

Schumacker, R. E., & Marcoulides, G. A. (1998). *Interaction and nonlinear effects in structural equation models*. Mahwah, NJ: Lawrence Erlbaum.

Stanne, M. B., Johnson, D. W., & Johnson, R. (1999). Does competition enhance or inhibit motor performance: A meta-analysis. *Psychological Bulletin, 125,* 133-154.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Tucker, L. R., & Lewis, C (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1-10.

Tufte, E. R. (2001). *The visual display of quantitative information (2nd ed.).* Cheshire, CT: Graphics Press.

Vroom, V. H. (1964). *Work and motivation.* New York: Wiley.

Yang Jonsson, F. (1997). *Non-linear structural equation models: Simulation studies of the Kenny-Judd model.* Acta Universitatis Upsaliensis Studia Statistica Upsaliensia 4 (Ph.D Dissertation, Uppsala University, Sweden).

Yang Jonsson, F. (1998). Modeling interactions and nonlinear effects: A step-by-step LISREL example. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation models* (pp. 17-42). Mahwah, NJ: Lawrence Erlbaum.

Yang-Wallentin, F., & Jöreskog, K. G. (2001). Robust standard errors and chi-squares for interaction models. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 159-171). Mahwah, NJ: Lawrence Erlbaum.

Yukl, G. (2002). *Leadership in organizations* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

APPENDIX A

ASSUMPTIONS AND STATISTICAL CONSTRAINTS

FOR THE JÖRESKOG-YANG METHOD

The Kenny-Judd model can be formulated in LISREL terminology as:

$$y = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta \tag{A.1}$$

The latent variables ξ_1 and ξ_2 are indicated by the observed variables x_1, x_2 and x_3, x_4 respectively. This relationship can be depicted:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{pmatrix} + \begin{pmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ 0 & \lambda_3 \\ 0 & \lambda_4 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{pmatrix} \tag{A.2}$$

Utilizing equation A.1 and A.2, the Jöreskog-Yang procedure rests on the following six assumptions (Jöreskog & Yang, 1996, p.58):

1. ξ_1 and ξ_2 are bivariate normal with zero means
2. $\zeta \sim N(0,\Psi)$
3. $\delta_i \sim N(0,\Theta_i)$, $i = 1,\ldots,4$
4. δ_i is independent of δ_j for $i = j$
5. δ_i is independent of ξ_j for $i = 1,\ldots,4$ and $j = 1,2$
6. ζ is independent of δ_i and ξ_j for $i = 1,\ldots,4$ and $j = 1,2$

In words, the assumptions are:

1. The latent predictors, ξ_1 and ξ_2 jointly follow a multivariate normal (i.e., bivariate normal) distribution. The latent variables are each centered about a mean of zero.
2. The residual in the structural equation, ζ, follows a normal distribution with a mean of zero, and a constant variance labeled Ψ.
3. The residuals of the indicators in each of the latent variables follow a normal distribution with a mean of zero and constant variance labeled Θ.
4. The residuals of the indicators in each of the latent variables are uncorrelated with one another.
5. The residuals of the indicators in each of the latent variable are uncorrelated with the latent variables.
6. The residual in the structural equation is uncorrelated with the residuals of the indicators of the latent variables and is uncorrelated with the latent variables.

The structural equation model (SEM) in equation A.1, is estimated, given the indicators from A.2 – x_1, x_2, x_3, x_4 – as a typical SEM, with x_1 and x_3 serving as reference indicators for ξ_1 and ξ_2 respectively and the product is indicated by the product of x_1x_3. In addition the procedure requires the following constraints to estimate the product term $\xi_1\xi_2$ (Jöreskog & Yang, 1996; specific ordering of constraints are drawn from Jaccard & Wan, 1996):

1. The mean of ξ_1 and ξ_2 are fixed at 0 by the Kappa matrix
2. The mean of the product, $\xi_1\xi_2$, is constrained to equal the covariance between ξ_1 and ξ_2. This involves the Kappa and Phi matrices.
3. The covariance of between ξ_1 and $\xi_1\xi_2$ and the covariance between ξ_2 and $\xi_1\xi_2$ are constrained to equal zero via the Phi matrix.
4. The variance of $\xi_1\xi_2$ is constrained to equal the variance of ξ_1 (φ_{11}) times the variance of ξ_2 (φ_{22}) plus the squared covariance between ξ_1 and ξ_2 (φ_{21}). This involves the Phi matrix.
5. α in equation A.1 is constrained to equal 0 for the model to be identified.
6. The measurement error variance for the product term indicator is constrained to equal:
$$\tau_1^{\,2}\theta_3 + \tau_3^{\,2}\theta_1 + \phi_{11}\theta_3 + \phi_{22}\theta_3 + \theta_1\theta_3 \tag{A.3}$$
where τ, θ, and φ are defined as in equation A.1, A.2 and the assumptions above.
7. The covariance between δ_1 and δ_5 (the error of the product indicator) is constrained to equal $\tau_3*\theta_1$ and the covariance between δ_3 and δ_5 is constrained to equal $\tau_1*\theta_3$. These constraints involve the Tau-X matrix and the Theta-delta matrix.
8. τ_5 (the intercept for the regression of $\xi_1\xi_2$ onto the product indicator, x_1x_3) is constrained to equal $\tau_1*\tau_3$. This constraint involves the Tau-X matrix.
9. The observed product indicator, x_1x_3, is influenced by the latent variables ξ_1 and ξ_2. As such, the path from ξ_1 to x_1x_3 (λ_{51}) is constrained to equal τ_3. The path from ξ_2 to x_1x_3 (λ_{52}) is constrained to equal τ_1. The path from $\xi_1\xi_2$ to x_1x_3 (λ_{53}) is fixed to 1.0.

APPENDIX B

ITEMS USED IN STUDY 1

Instrumental coworker support ($\alpha = .92$)

1. Your coworkers would fill in while you're absent.
2. Your coworkers are helpful in getting job done.
3. Your coworkers give useful advice on job problems.
4. Your coworkers assist with unusual work problems.
5. Your coworkers will pitch in and help.

Affective coworker support ($\alpha = .92$)

1. Your coworkers really care about you.
2. You feel close to your coworkers.
3. Your coworkers take a personal interest in you.
4. You feel appreciated by your coworkers.
5. Your coworkers are friendly to you.

Competitive psychological climate ($\alpha = .77$)

1. My manager frequently compares my performance with that of my coworkers.
2. The amount of recognition you get in this company depends on how you perform compared to others.
3. Everybody is concerned with being the top performer.
4. My coworkers frequently compare their performance with mine.

Trait competitiveness ($\alpha = .88$)

1. I enjoy working in situations involving competition.
2. It is important to me to perform better than others on a task.
3. I feel that winning is important in both work and games.
4. I try harder when I am in competition with other people.

APPENDIX C

LISREL CODE FOR MODEL ESTIMATION IN STUDY 1

```
/* Example LISREL syntax for the estimation of a multiplicative model when component
terms are measured with four indicators */

JY method using mean centered items as indicators
SY=ITEMC.DSF
mo ny=5 nx=9 ne=1 nk=3 td=sy,fi ga=fr tx=fr ka=fr al=fr ty=fr
se
1 2 3 4 5 9 6 7 8 12 10 11 13 14/
fi lx 1 1 lx 5 2 lx 9 3
va 1.0 lx 1 1 lx 5 2 lx 9 3
fr td 1 1 td 2 2 td 3 3 td 4 4
fr td 5 5 td 6 6 td 7 7 td 8 8
fr ly 1 1 ly 2 1 ly 3 1 ly 4 1 ly 5 1
fr lx 2 1 lx 3 1 lx 4 1
fr lx 6 2 lx 7 2 lx 8 2
fi ka 1 ka 2                                    !constraint # 1
co ka 3 = ph 2 1                                !constraint # 2
fi ph 3 1 ph 3 2                                !constraint # 3
co ph 3 3 = ph 1 1*ph 2 2 + ph 2 1*ph 2 1       !constraint # 4
fi al 1                                         !constraint # 5
                                                !constraint # 6
co td 9 9 = tx 1**2*td 5 5 + tx 5**2*td 1 1 + c
        ph 1 1*td 5 5 + ph 2 2*td 1 1 + td 1 1*td 5 5
                                                !constraint # 7
co td 9 1 = tx 5*td 1 1
co td 9 5 = tx 1*td 5 5
co tx 9 = tx1*tx5                               !constraint # 8
                                                !constraint # 9
co lx 9 1 = tx 5
co lx 9 2 = tx 1
ou ad=off it=500 nd=4
```

APPENDIX D

PRELIS PROGRAM FOR DATA GENERATION FOR STUDY 2

```
Create Variables
da no=800 rp = 500
co all

!Generate latent exogenous variables, r_xz = .3
NE V1 = NRAND; NE V2 = NRAND; NE V3 = NRAND
NE X = V1; NE Z = .3*V1+.9539392*V2
NE XZ = X*Z

!Structural equation for estimation of latent endogenous variable
 NE Y = -.25*X+.12*Z+.125*XZ + V3       !f^2 = .01 – set random seed = 12345
! NE Y = -.25*X+.12*Z+.150*XZ + V3      !f^2 = .02 – set random seed = 54321
! NE Y = -.25*X+.12*Z+.400*XZ + V3      !f^2 = .15 – set random seed = 34567

!Generate random errors for manifest indicators
NE TE1 = .5*NRAND; NE TE2 = .5*NRAND; NE TE3 = .5*NRAND

NE TD1 = .866*NRAND; NE TD2 = .935*NRAND; NE TD3 = 1.0*NRAND
NE TD4 = 1.061*NRAND; NE TD5 = 1.118*NRAND; NE TD6 = 1.173*NRAND
NE TD7 = 1.225*NRAND; NE TD8 = 1.275*NRAND; NE TD9 = 1.323*NRAND
NE TD10 = 1.369*NRAND; NE TD11 = 1.414*NRAND; NE TD12 = 1.458*NRAND

NE TD13 = .866*NRAND; NE TD14 = .935*NRAND; NE TD15 = 1.0*NRAND
NE TD16 = 1.061*NRAND; NE TD17 = 1.118*NRAND; NE TD18 = 1.173*NRAND
NE TD19 = 1.225*NRAND; NE TD20 = 1.275*NRAND; NE TD21 = 1.323*NRAND
NE TD22 = 1.369*NRAND; NE TD23 = 1.414*NRAND; NE TD24 = 1.458*NRAND

!Generate manifest indicators for endogenous variable
NE Y1 = 1*Y + TE1; NE Y2 = .95*Y + TE2; NE Y3 = .90*Y + TE3

!Generate manifest indicators for exogenous variables
NE X1 = .95*X + TD1; NE X2 = .925*X + TD2; NE X3 = .9*X + TD3
NE X4 = .875*X + TD4; NE X5 = .85*X + TD5; NE X6 = .825*X + TD6
NE X7 = .8*X + TD7; NE X8 = .775*X + TD8; NE X9 = .75*X + TD9
NE X10 = .725*X + TD10; NE X11 = .7*X + TD11; NE X12 = .675*X + TD12

NE Z1 = .95*Z + TD13; NE Z2 = .925*Z + TD14; NE Z3 = .9*Z + TD15
NE Z4 = .875*Z + TD16; NE Z5 = .85*Z + TD17; NE Z6 = .825*Z + TD18
NE Z7 = .8*Z + TD19; NE Z8 = .775*Z + TD20; NE Z9 = .75*Z + TD21
NE Z10 = .725*Z + TD22; NE Z11 = .7*Z + TD23; NE Z12 = .675*Z + TD24
```

```
!Create single product indicator for item level
!NE X1Z1 = X1*Z1

!Parcels are created according to study design: congeneric vs. balanced and item/parcel
!ratios An example for congeneric for 3 parcels with 4 items follows

!Create congeneric parcels
!NE D = 4**-1
!NE CPX1 = X1+X2+X3+X4
!NE CPX1 = CPX1*D
!NE CPX2 = X5+X6+X7+X8
!NE CPX2 = CPX2*D
!NE CPX3 = X9+X10+X11+X12
!NE CPX3 = CPX3*D

!NE CPZ1 = Z1+Z2+Z3+Z4
!NE CPZ1 = CPZ1*D
!NE CPZ2 = Z5+Z6+Z7+Z8
!NE CPZ2 = CPZ2*D
!NE CPZ3 = Z9+Z10+Z11+Z12
!NE CPZ3 = CPZ3*D

!NE CPX1Z1 = CPX1*CPZ1

!Select and delete temporary variables
SD V1-V3 Y X Z XZ TE1-TE3 TD1-TD24
OU ma=cm RA=ITEM.DAT IX=12345
```

VITA

Thomas D. Fletcher

EDUCATION

2005 Old Dominion University
Ph.D., Industrial/Organizational Psychology

2002 Old Dominion University
M.S., Psychology

1999 Georgia State University
B.S., Psychology; minor: Political Science
Magna Cum Laude

PUBLICATIONS

Ehler, M. L., Major, D. A., & Fletcher, T. D. (2003). Applying I-O to medicine: Making the case that it can be done and should be done. *The Industrial-Organizational Psychologist, 41*, 50-54. *Non-refereed publication.

Fletcher, T. D. & Major, D. A. (2004). Medical students motivations' to volunteer: Examining the nature of gender differences. *Sex Roles, 51,* 109-114.

Major, D. A., Turner, J. E., & Fletcher, T. D. (in press). Personality predictors of motivation to learn: An Examination of the Big Five and Proactive Personality. *Journal of Applied Psychology.*

TEACHING EXPERIENCE

Instructor, 8/04 – 5/05, Motivation. Department of Psychology, Old Dominion University. Developed lecture and taught course. The course is 300 level and covers the psychological theories of motivation including the role of needs, cognitions, emotions, and individual differences in motivated behavior.

Adjunct Lecturer, 8/03 to 5/04, Tests and Measurements. Department of Psychology, Christopher Newport University. Developed lecture, taught course and conducted two labs for each hour of lecture to undergraduates. The course is 400 level and covers theory and application of psychometric principles.

Co-Instructor, 1/03 to 5/03, Advanced Personnel Psychology II. Department of Psychology, Old Dominion University. Developed and disseminated lecture on multilevel modeling (6 weeks).